

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2002, 12:10:12 ; Search time 57.07 Seconds  
(without alignments)  
903.071 Million cell updates/sec

Title: US-09-863-824-2

Perfect score: 2517

Sequence: 1 MVRLAAELLLLGLLLTLH.....CTESPSDEYIKQFQAREAY 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_032802.\*
- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
  - 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
  - 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
  - 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
  - 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
  - 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
  - 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
  - 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
  - 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
  - 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
  - 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1353	53.8	246	22	AA25462 Human protein sequ
2	1035	41.1	198	21	AB41922 Human ORFX ORF1686
3	927	36.8	571	22	AB88393 Human membrane or
4	927	36.8	571	22	AB49765 Human proliferatio
5	836	33.2	149	22	ABG06346 Novel human diagno
6	315	12.5	215	22	ABG06347 Novel human diagno
7	225.5	9.0	914	22	ABG27398 Novel human diagno
8	182	7.2	156	22	AAU21208 Human novel foetal
9	143	5.7	1572	20	AAW99301 Human BAI2 protein
10	137.5	5.5	1584	20	AAW99300 Human BAI1 protein
11	134.5	5.3	1522	20	AAW99302 Human BAI3 protein

12	134.5	5.3	1522	21	AA23601 Human secreted pr
13	133.5	5.3	763	22	AB62815 Drosophila melandog
14	131.5	5.2	1934	22	AB72301 Human ADAMTS-9 alt
15	130	5.2	1328	22	ABG22373 Novel human diagno
16	129	5.1	1206	19	AAW47030 Bovine N-proteinase
17	128.5	5.1	242	21	AA800040 Human COMP/TSP-1 c
18	128	5.1	239	14	AA40823 Human thrombospond
19	128	5.1	1211	19	AAW47028 Human N-proteinase
20	127	5.0	57	20	AA49505 Human METH1 thombo
21	127	5.0	57	22	AA50007 TSPI domain #3. H
22	127	5.0	555	22	AAU2914 Angiotensin conver
23	127	5.0	731	22	AAU02913 Angiotensin conver
24	127	5.0	1152	21	AA800042 Human thrombospond
25	127	5.0	1170	22	AA90800 Human shear stress
26	127	5.0	1170	22	AA874450 Human variant thro
27	126	5.0	206	21	AA808135 Amino acid sequenc
28	125.5	5.0	873	22	AB66441 Drosophila melandog
29	125	5.0	4561	22	ABG30203 Novel human diagno
30	125	5.0	9222	22	ABG21064 Novel human diagno
31	124	4.9	969	21	AA53900 Amino acid sequenc
32	124	4.9	1048	22	AB85695 Larynx carcinoma a
33	124	4.9	1203	22	AB50004 Bovine metalloprot
34	124	4.9	2150	21	AA53898 Amino acid sequenc
35	124	4.9	2165	22	AA90617 Human secreted pro
36	123	4.9	157	21	AA808133 Amino acid sequenc
37	122	4.8	837	20	AAW75425 Human aggrecan deg
38	122	4.8	837	21	AA99429 Human PRO1563 (UNQ
39	122	4.8	837	22	AA878228 Human aggrecanase-
40	122	4.8	837	22	AAU29199 Human PRO polypept
41	122	4.8	837	22	AA866178 Protein of the inv
42	122	4.8	840	21	AB21256 Human metalloprote
43	121.5	4.8	1882	22	AA872286 Human ADAMTS-9 am
44	120.5	4.8	459	22	AAU2916 Angiotensin conver
45	118.5	4.7	1088	22	AA667244 Amino acid sequenc

ALIGNMENTS

RESULT 1  
AA25462  
ID AA25462 standard; Protein; 246 AA.  
AC AA25462;  
DT 16-OCT-2001 (first entry)  
XX Human protein sequence SEQ ID NO:977.  
DE  
XX Human; cancer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggagant; haemostatic; vulnery; antidiabetic; eczema;  
KW dermatologic; antiallergic; antidiabetic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.

Homo sapiens.

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US35017.

PR	23-DEC-1999;	99US-0471275.	
PR	21-JAN-2000;	2000US-0488725.	
PR	25-APR-2000;	2000US-0552317.	
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Drmanac RT;		
FI	WPI; 2001-457603/49.		
DR	N-PSDB; AAH99403.		
XX	Isolated human polynucleotides encoding polypeptides, useful for the		
PT	treatment and diagnosis of e.g. cancer, ulcers and HIV infection -		
XX	Claim 20; Page 206; 1217pp; English.		
XX	AAH99166 to AAH99904 encode the human proteins given in AAH25225 to		
CC	AAH25963. The proteins can have activities based on the tissues and		
CC	cells they are expressed in, such as: antiinflammatory; antirheumatic;		
CC	antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;		
CC	central nervous system; virucide; anti-HIV; fungicide; antimutagen;		
CC	cardiovascular; antianemic; antiagregant; haemostatic; vulnerary;		
CC	antitumor; osteopathic; dermatological; antiallergic; antilesthetic;		
CC	antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;		
CC	antiparkinsonian; and immunostimulant. The proteins and polynucleotides		
CC	encoding them can be used in gene therapy, antisense therapy and vaccine		
CC	production, the proteins and polynucleotides are useful for screening for		
CC	agonists or antagonists of a protein and for the treatment and diagnosis		
CC	of disorders associated with the activity of a protein e.g. inflammation,		
CC	rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,		
CC	neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal		
CC	infections, autoimmunity, genetic diseases, haematopoietic disorders,		
CC	anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,		
CC	osteoporosis, severe combined immunodeficiency, eczema, allergic		
CC	rinitis, asthma, diabetes, cancer, multiple sclerosis, depression,		
CC	Alzheimer's disease, Parkinson's disease, neurodegenerative and		
CC	neurological disorders.		
XX	Sequence 246 AA;		
SQ			
Query Match 53.8%; Score 1353; DB 22; Length 246;			
Best Local Similarity 99.6%; Pred. No. 1.5e-120;			
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	175	DQDYXXVDSNDDSNFLNPPRGWDHTAPGHRFTETKDQPEYDSTDGEGWLSVSGSVTCG	234
Db	1	dqdykdydstddsnfnpprgwdhtapghrtfctkdqpeydstdgewslwsvcsvtcg	60
QY	235	NGNQKTRSCGYACTATESRCDRPNCPGIEDTFFTAATEVSLLAGSEEFNATKLFVDVT	294
Db	61	ngnqktrscgyactatesrctdrpncpgiedtffrtaatevslagseefnatklfevdt	120
QY	295	DSCEWRMSCKSEFLKYYMHVKNWDLSPSCSPYTEVAYSTADIEDRKDFRWKDSGP	354
Db	121	dscerwmsckseflkyyhmvmndlpsscspytevaystadifdrkdrfwkdsagp	180
QY	355	KEKLEIYKPARVCIRSMLSLESTTLAAOHCYGDNNQLITRGKAGTGNLSTFEPSAEL	414
Db	181	kekleykparvcirsmlesleltlaaahccygdnnqmlitrgkgagtgnllestfesa	240
QY	415	HYKVDV 420	
Db	241	hykvdv 246	
RESULT 2			
AAB41922			
ID	AAB41922 standard; Protein: 198 AA.		
XX			
AC	AAB41922;		
XX			
DT	08-FEB-2001 (first entry)		
XX			

XX	Human OREF ORF1686 polypeptide sequence SEQ ID NO:3372.		
DE	Human; open reading frame; OREF; detection; cytotatic; hepatotropic;		
XX	vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	hypotensive; dermatological; immunosuppressive; antileukemic;		
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;		
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	neurodegenerative disorder; osteoarthritis; hypothyroidism; SCID; AIDS;		
KW	cardiovascular disease; diabetes mellitus; graft vs host disease;		
KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antileukemic disease; coagulation;		
KW	thrombosis; contraceptive.		
XX	Homo sapiens.		
OS	WO2000058473-A2.		
XX	05-OCT-2000.		
PN	31-MAR-2000; 2000WO-US08621.		
PD	31-MAR-1999; 99US-0127607.		
XX	02-APR-1999; 99US-0127636.		
XX	05-APR-1999; 99US-0127728.		
XX	30-MAR-2000; 2000US-0540763.		
XX	(CURA-) CURAGEN CORP.		
XX	Shimkets RA, Leach M;		
PI	WPI: 2000-602362/57.		
DR	N-PSDB; AAC76131.		
XX	Novel nucleic acids and peptides derived from open reading frame X,		
PT	useful for treating e.g. cancers, proliferative disorders,		
PT	neurodegenerative disorders and cardiovascular disease -		
XX	Claim 11; Page 2553-2554; 5507pp; English.		
XX	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,		
CC	which represent the human OREF open reading frames 1 to 3161. The OREF		
CC	sequences have activities such as: cytotatic; hepatotropic; vulnerary;		
CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;		
CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressive;		
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;		
CC	antidiabetic; hypotensive; dermatological; antirheumatic;		
CC	antileukemic; antibacterial; antiviral; antifungal; antirheumatic;		
CC	antithyroid; and antianemic. The sequences can be used for determining		
CC	the presence of or predisposition to, or preventing or treating		
CC	pathological conditions associated with an OREF-associated disorder. The		
CC	nucleic acids can be used to express OREF proteins in gene therapy		
CC	vectors. The proteins and nucleic acids may be used to treat cancers,		
CC	proliferative disorders, neurodegenerative disorders, diabetes mellitus,		
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,		
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus		
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,		
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,		
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,		
CC	nocturnal haemoglobinuria, antileukemic disease; to enhance		
CC	coagulation; to inhibit thrombosis; and as a contraceptive.		
XX	Sequence 198 AA;		
SQ			
Query Match 41.1%; Score 1035; DB 21; Length 198;			
Best Local Similarity 97.3%; Pred. No. 2.4e-90;			
Matches 183; Conservative 3; Mismatches 2; Indels 0; Gaps 0;			

Qy 60 SFSLKSEAPREHLQHQAHPFPRFRQETGHSFLQDRFSLFLDLPNFPDLKADIN 119  
 Db 11 afalakeaprehldhqaahqfprfrfqtghpslqrdfsrflldlpnfpdlksadin 70  
 Qy 120 COXENIQVTIEVDPGSEADKQHPENKPSWSVSPDWRAMWORSLSLARANGDQDYX 179  
 Db 71 gqnpniqvievdpdseadkqhpknkpswsvpspdwrawwgrsrlsrlansgdqdyk 130  
 Qy 180 YDSTSDSNFLNPPRGWDHTAPGHRTPETKQDPEYDSDGSDWSLWSVSVTCGNGNOK 239  
 Db 131 ydstsdnflnpprgwdhtapghrtetkqdpdydstdgdsdswslwsvsvtcgngnqk 190  
 Qy 240 RTRSCGYA 247  
 Db 191 rtrscgya 198

RESULT 3  
 AAB88393  
 ID AAB88393 standard; Protein; 571 AA.  
 XX  
 AC AAB88393;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX  
 DE Human membrane or secretory protein clone PSEC0137.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-0114090.  
 XX  
 PR 08-JUL-1999; 99JP-0194179.  
 PR 11-JAN-2000; 2000JP-0118775.  
 PR 02-MAY-2000; 2000JP-0183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX  
 DR WPI; 2001-093989/11.  
 DR N-PSDB; AAF93820.  
 XX  
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development -  
 XX  
 PS Claim 1; SEQ ID 154; 609pp + CD ROM; English.

XX  
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by  
 CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic

CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.  
 XX  
 SQ Sequence 571 AA;  
 Query Match 36.8%; Score 927; DB 22; Length 571;  
 Best Local Similarity 42.7%; Pred. No. 2.3e-79;  
 Matches 200; Conservative 62; Mismatches 130; Indels 76; Gaps 17;  
 Qy 60 SFSLKSEAPREHLQHQAHPFPRFRQETG-----HPSLQRDF----- 99  
 Db 117 stlistpnp-----dtqasaspdprrleeeearllprthlqaelhqhcwtvtepaalt 172  
 Qy 100 -----PRS-----FLDLNPNFDLSKADINQXPNQVTVIEVDPGPDSEADK--QHPE 147  
 Db 173 gnatpprtqevtpilllelqkipeivhatlstpnqndqvtikvvedpqaevidllaepsn 232  
 Qy 148 KP-----SWSVSPDWRAMW-----QRSLSLARANGS-----DQDY-----XYDSTS 184  
 Db 233 pppqdtlsw-lpal-wpflwgdykgeekdrapgekeedepysediegedqdekee 290  
 Qy 185 DDSN-----FLNPPRGWDH--TAPGHRTPETKQDPEYDSDGSDWSLWSVSVTCGNGNQ 238  
 Db 291 deeeqalwfnottndwdqglapgdwvf--kdsvsyd-yepqkewspwpcsgncstqkq 347  
 Qy 239 KTRSCGYACTATESRTCDRPNCPGIE--DTFRTAATVSLLAGSEEFNATKLFVDFDTS 296  
 Db 348 qtrpcgygctatetrtcdlpscpgetdkdtlgipseeekllar----natdmhdqdvds 403  
 Qy 297 CERWMSKSEFLKYMVKVMNDLPCSCPSYPTVAYSTADIFDRKRDPRKWDASGPKE 356  
 Db 404 cekwlncsksfliksqmlrdlpscpaypleamdspsvsgdehqgfrsfrwrdasgpre 463  
 Qy 357 KLEIYKPTARYCIRSMLESLESTTLAAQHCCYGDNNMOLITRGKAGCTPNLISTEFAELHY 416  
 Db 464 rldlyqptarfclrmslsgesstlaahccydederslltrkgagampnlstdfspklnhf 523  
 Qy 417 KVDVLPWICKGWSRYNEARPPNNGOKCTESPSEDYIKQFQEAERY 464  
 Db 524 kfdttplwlckgdwsrlhavlpnngnactdnpnpleeeylaqlqaekey 571  
 RESULT 4  
 AAB49765  
 ID AAB49765 standard; Protein; 571 AA.  
 XX  
 AC AAB49765;  
 XX  
 DT 20-APR-2001 (first entry)  
 XX  
 DE Human proliferation differentiation factor amino acid sequence.  
 KW Human; proliferation differentiation factor; haematopoietic function.  
 OS Homo sapiens.  
 PN WO200104312-A1.  
 XX  
 PD 18-JAN-2001.  
 XX  
 PF 06-JUL-2000; 2000WO-JP04514.  
 XX  
 PR 08-JUL-1999; 99JP-0194179.  
 PR 18-OCT-1999; 99US-0159586.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;  
 XX

DR WPI: 2001-138354/14.  
DR N-PSDB; AAF29348.  
XX proliferation-differentiation factor protein encoded by PSEC137 cloned  
XX from human cDNA library, being hematopoietic factor for inducing  
PT differentiation of blood cells, used to maintain hematopoietic function  
PT -  
XX  
XX Claim 1; Page 42-44; 49pp; Japanese.  
XX  
XX This invention relates to polynucleotide sequence PSEC137, which encodes  
CC a human proliferation differentiation factor protein. Included in the  
CC invention is a vector containing the cDNA sequence, a transformant  
CC containing the vector, and a process for producing the protein. The  
CC protein together with its encoded DNA and antisense DNA are used in drugs  
CC to maintain the haematopoietic function. The present sequence represents  
CC the human proliferation differentiation factor protein amino acid  
CC sequence.  
XX Sequence 571 AA;  
SQ

Query Match 36.8%; Score 927; DB 22; Length 571;  
Best Local Similarity 42.7%; Pred. No. 2.3e-79;  
Matches 200; Conservative 62; Mismatches 130; Indels 76; Gaps 17;  
QY 60 SFSLSKEAPREHLHQAAHQPFRFRQETG----HPSLORDF----- 99  
Db 117 stlslatpnp-----dtqasapdrprlreecearllprthlqaelhghcwtvtepaalt 172  
QY 100 -----PRSL-----FLIDLFPNFDLSKADINGQXPNQIVTIEVDPGPDSEADKD--QUPEN 147  
Db 173 gnatprtqevtpllleiklpelvatlstpnqnvtklvvedpqaevsidllaepsn 232  
QY 148 KP-----SNVSPDPRAWW-----ORSLSLARANS-----DODY-----XYDSTS 184  
Db 233 pppqdtlsw-lpal-wpflwgdykgeekdrapgekgeekeddypsedlegedgedkee 290  
QY 185 DDSN-----FLNPPRGWDH--TAPGHRFTETKDQPEYDSTDGDSWLSVCSVTCGNQ 238  
Db 291 deeeqalwngttdndwqglapdwf--kdsvsyd-yepqkewspwpcsgncstgkq 347  
QY 239 KTRSCGYACTATESRTCDRPNCGIE--DTFRTAATEVSLLAGSEEFNATKLFVEVDTS 296  
Db 348 qrtprcgygtatetrtcdlpscpgetdkdtlglpseewkllar----natdmhdqdvds 403  
QY 297 CERWMSCKSEFLKKYMHKYNLDLPCSPCYPTPEVAYSTADIFDTRKDRWKDASGPKE 356  
Db 404 cekwlncksdflkylsqmldrldpspcpcaypleamdpsvlgdehqgrsfrwrdasgpre 463  
QY 357 KLEYKPTARYCIRSMLSLESTTLAAQHCCYGDNMOLITRGAGCTPNLITFEFSAELHY 416  
Db 464 rldiyptarfclrmslsgesstlaahccydedrslltrgkagmpnlilstdfspklhf 523  
QY 417 KVDVLPWIIKGDWSRYNEARPNPNCQKCTESPDDEYIKQFQEARAY 464  
Db 524 kfdtptwilckgdwsrlhavlppnngactdnplpseeiaqlqaekey 571

RESULT 5  
ABG06346  
ID ABG06346 standard; Protein; 149 AA.  
XX  
XX ABG06346;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #6337.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
KW  
XX Homo sapiens.  
OS

XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
PF 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
XX N-PSDB; AAS70533.  
DR New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PT  
XX Claim 20; SEQ ID No 36705; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC a food supplement. (II) and its binding partners are useful in medical  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 149 AA;  
Query Match 33.2%; Score 836; DB 22; Length 149;  
Best Local Similarity 99.3%; Pred. No. 1.5e-71;  
Matches 148; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 175 DDYXYDSTSDSFLNPPRGWDHTAGCHRTFFETKDDQPEYDSTDGDSWLSVCSVTCG 234  
Db 1 ddykydstsdssnflnpprgwdhtapghrtfctkddpeydstdgedswlsvcsvtcg 60  
QY 235 NGNOKRTBSCGYACTATESRTCDRPNCGIEDTFRTAATEVSLLAGSEEFNATKLFVEVD 294  
Db 61 ngngkrtbscgyactatesrtcdprncpgiedtfrrtaatevslagseefnatklfevdt 120  
QY 295 DSCERWMSCKSEFLKKYMHKYNLDLPCSP 323  
Db 121 dscerwmsckseflkkyhmhkvmdlpscp 149  
RESULT 6  
ABG06347  
ID ABG06347 standard; Protein; 215 AA.  
XX  
XX ABG06347;  
XX  
XX 13-FEB-2002 (first entry)  
DT



[illegible]

	l;   l;l:
Db	172 slaglslnqgrscwge 188
RESULT	7
ABG27398	
ID	ABG27398 standard; Protein; 914 AA.
XX	AC
AC	ABG27398;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #27389.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
KW	
OS	Homo sapiens.
XX	
XX	WO200175067-A2.
PN	XX
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
PT	(HYSE-) HYSEQ INC.
PA	
XX	Dmanac RT, Liu C, Tang YT;
PI	XX
XX	WPI; 2001-639362/73.
DR	N-PSDB; AAS91585.
DR	
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutation
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
PT	
PS	Claim 20; SEQ ID No 57757; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probe
CC	polymerase chain reaction (PCR) primers, oligomers, and for chro
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequen
CC	for identifying expressed genes. (I) is useful in gene therapy t
CC	to restore normal activity of (II) or to treat disease states i
CC	(II). (II) is useful for generating antibodies against it, detecte
CC	quantitating a polypeptide in tissue, as molecular weight marker
CC	a food supplement. (II) and its binding partners are useful in m
CC	imaging of sites expressing (II). (I) and (II) are useful for th
CC	disorders involving aberrant protein expression or biological ac
CC	The polypeptide and polynucleotide sequences have applications i
CC	diagnostics, forensics, gene mapping, identification of mutation
CC	responsible for genetic disorders or other traits to assess bloc
CC	and to produce other types of data and products dependent on DNA
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the pr
CC	specification, but was obtained in electronic format directly fr
CC	at ftp.wipo.int/pub/published_pct_sequences.

Query Match 9.0%; Score 225.5; DB 22; Length 914;  
Best Local Similarity 87.9%; Pred. No. 3.7e-12;

QY 1 MVRLAEEILLGLLLTLHITVLRGSAADGPDAAAGNASQAQLQNINLVGSDTTSE 58

Fri Apr 26 10:29:58 2002

us-09-863-824-2.rag

Db 487 mvrtaealllllllllthitvrlsgaadgpdaaagaaqqlq-glllgkptlse 543

RESULT 8

AAU21208  
ID AAU21208 standard; Protein; 156 AA.

XX AC AAU21208;

XX 18-DEC-2001 (first entry)

XX Human novel foetal antigen, SEQ ID NO 1452.

XX Human; foetal tissue antigen; antiinflammatory; neuroprotective;  
KW immunomodulator; cardiovascular; cytostatic; nephroprotectic;  
KW cardiovascular; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; breast neoplasm; cancer;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.

XX Homo sapiens.

XX WO200155312-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01321.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.



Db	577	shevrylylsrehlagqrlagegmsqvr-----slqellarrtyysgdllifsvdil- 631	
Qy	423	WIICKGWSRYNEARPPNNGKCTEPSPDEYIKQFQ 459	
Db	632	-----rntvdtfkratyvpsaddvqrfq 655	
RESULT 10			
ID	AAW99300	AAW99300 standard; Protein; 1584 AA.	
XX	AC	AAW99300;	
XX	DT	21-MAY-1999 (first entry)	
XX	DE	Human BAI1 protein.	
XX	KW	Human; BAI1; brain; cancer; drug; diagnosis; prevention; treatment.	
XX	OS	Homo sapiens.	
XX	PN	JP11032766-A.	
XX	PD	09-FEB-1999.	
XX	PF	16-JUN-1997; 97JP-0176485.	
XX	PR	23-MAY-1997; 97JP-0150460.	
XX	PA	(SAKA ) OTSUKA PHARM CO LTD.	
XX	WPI	1999-183823/16.	
DR	N-PSDB	AAAX21355.	
XX	PT	New human BAI gene - is expressed in brain plays important role in cancer formation	
XX	PS	Claim 1; Page 18-22; 62pp; Japanese.	
XX	CC	This sequence represents the human BAI1 protein. The gene is expressed specifically in the brain and plays an important role in cancer formation in the brain. The BAI1 protein can be used in drug compositions to diagnose, prevent or treat such cancers.	
XX	SQ	Sequence 1584 AA;	
Query Match 5.5%; Score 137.5; DB 20; Length 1584;			
Best Local Similarity 21.8%; Pred. No. 0.0021;			
Matches 119; Conservative 50; Mismatches 177; Indels 201; Gaps 30;			
Qy	2	VRLAELLGLLGLLTLHTVLRGSGADGPDAA-----GNASQAQLNNLNVGS 53	
Db	11	vvilaplllll--llllgraraagadagpcepcaatlvggkffgfsaaa-----vfp 62	
Qy	54	DTTSETSFSLSKAPREH-LDHOAAHQPPRPFRQETCHPSLQRDFP-RSFL-----L 105	
Db	63	anasrcswtlrnpdrrtytlmkvakapvp-----csgpgrvtyqfdfslestrtyl 115	
Qy	106	DLNFPD-LSKADINGQXPNQIOTIEVDPDSEADKDQHPEN---KPSWSVPSP--DW- 158	
Db	116	gvesfdevlrlcdpsaplaflgaskqfl-----gmrrgppqhdglrpragppgptddfs 170	
Qy	159	-----RA-----WQRSLSLARAN-----SGDQDYXYDSTSDD 186	
Db	171	veylvgnrnpstraacqmlorwidaclagsrshpccimgimtpcaclg9geag----- 221	
Qy	187	SNFLNP-----PRG-----WDHTAPGHR---TFETKDQPEYDSTGDGWSLWSVCVTC 233	
Db	222	-----gpaagplargdvclrdavaggpenciltsitdrgghgatggkwlswlwgectrdc 277	
Qy	234	GNGNQKTRSC-----GYACTAT--ESRTCDRPNCPGIEDIFRTAAATEVSLLA----- 279	

Db	278	ggglqtrtrtclpapvgveggcgvleegrqcnreac---gpagrtssrsqslrstddarr 334	
Qy	280	-----GSE-----EFNATKLEFVDTDCERWMSCKSEFLKYYMHKVMNDLPSCPCSYPTVA 331	
Db	335	reelgdclqfgfpapqtdgpaeeewspwsvcss-----tcg----- 371	
Qy	332	YSTADIFDRIKRRKDFRKDASGPKKELEYKPTARYCIRSMLSLE-STTLAAQHCCYGDN 390	
Db	372	-----egqtrtrfcvssysystgcsgplreqrlc---n 401	
Qy	391	MLITRGKAGTPNLISTEFSAEHLHYKVDVLPWIIIC-----KGDWSRYNEARPPNNGKC 445	
Db	402	nsavcpvhgaw-----dews-----pwslcstcggrfdrtrtcrppqfggnp 445	
Qy	446	TESPSDE 452	
Db	446	cegpekq 452	

RESULT 11

AAW99302

ID AAW99302 standard; Protein; 1522 AA.

XX AC AAW99302;

XX DT 21-MAY-1999 (first entry)

XX DE Human BAI3 protein.

XX KW Human; BAI3; brain; cancer; drug; diagnosis; prevention; treatment.

XX OS Homo sapiens.

XX PN JP11032766-A.

XX PD 09-FEB-1999.

XX PF 16-JUN-1997; 97JP-0176485.

XX PR 23-MAY-1997; 97JP-0150460.

XX PA (SAKA ) OTSUKA PHARM CO LTD.

XX WPI; 1999-183823/16.

XX DR N-PSDB; AAX21357.

XX PT New human BAI gene - is expressed in brain plays important role in cancer formation

XX PS Claim 2; Page 26-30; 62pp; Japanese.

XX CC This sequence represents the human BAI3 protein. The gene is expressed specifically in the brain and plays an important role in cancer formation in the brain. The BAI3 protein can be used in drug compositions to diagnose, prevent or treat such cancers.

XX SQ Sequence 1522 AA;

Query Match 5.3%; Score 134.5; DB 20; Length 1522;

Best Local Similarity 21.4%; Pred. No. 0.0038;

Matches 84; Conservative 48; Mismatches 170; Indels 91; Gaps 21;

Qy 62 SLUSKEA---PREHL-----DHQAAHQPPRPFRQETGHPSLQRDFPRSFLLD-----LPN 109

Db 241 nltreakppkeefmgmgdhtiksq---rprsvhekrypqqadaak-fmaqtgesvvee 296

Qy 110 FFDLSKADIN-CQXPNIQVTIEV-----VDGPDSEADKDQHPENKPSWSVPSPDRAWW 162

Db 297 wsgwtatcsvtcgsgsqvrrtctcvsygtchsgplresrvcmntalcphvgv-----weews 352

Qy 163 QRSLSLARANSGDQDYXYDSTSDDSNFLNPPRGWDHTAPGHRFTFETKDQPEYDSTDGE-G 221

```
Db 353 pwslcstcgrgtrtrsc-----ppqyggprcpgpethhhkpcnialcpvdgqw 404
QY 222 DWSLWSVCSVTCGNGNQKTRSC-----GYACTA--TESRTCDRPNCPGIEDTFTTAAT 273
Db 405 ewsswqscvtsngtqgrsrqtaahgscrgpwaesrecynpec-----452
QY 274 EVSLLAGSEEFNATKLFVDTSCE-----RWMSCKSEFLK-KYMHKVMNDLPSCP--C 324
Db 453 -----tangqwnqghwscskscdggwerrirtcgavitgqccgsevrrecseqrc 507
QY 325 SYPTEVA---YSTADIFDRKRDFRWK-----DASGPKEK---LEIY-----KPTARYC 368
Db 508 papyeicpedylnsmvwkrtpgdlafnqcpinatgttsrrcslnhgvafwqpsfarc 567
QY 369 IRS-----MLSLESTTLAAQHCCYGDNMQLITR 396
Db 568 lsneyrhlqhsikehlaqgrmlagdgmsqvtk 600

RESULT 12
ID AAB23601 standard; Protein; 1522 AA.
AC AAB23601;
XX
DT 12-JAN-2001 (first entry)
DE Human secreted protein SEQ ID NO: 2.
KW Human; secreted protein; cytokine; cell proliferation;
KW nutritional supplement; immune modulation; autoimmune disorder;
KW haematopoiesis regulation; tissue growth; haemostasis; inflammation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 3..15
FT Protein /label= signal_peptide
FT Protein 16..1522 /label= mature_protein
XX
PN WO200049134-A1.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04340.
XX
PR 19-FEB-1999; 99US-0120680.
PR 23-APR-1999; 99US-0298733.
PR 17-AUG-1999; 99US-0149639.
PR 23-SEP-1999; 99US-0155686.
PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167822.
PR 29-NOV-1999; 99US-0167823.
PR 15-FEB-2000; 2000US-0298733.
XX
PA (ALPH-) ALPHAGENE INC.
XX
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Repiejko P;
XX
DR WPI; 2000-549267/50.
DR N-PSDB; AAA93101.
XX
PT New secreted proteins and polynucleotides encoding them, which are
PT derived from Homo sapiens, useful for therapy, diagnosis, and research,
PT as well as nutritional sources or supplements -
XX
PS Claim 9; Page 234-238; 309pp; English.
XX
CC The present sequence is the sequence of a human secreted protein. Its
CC cDNA was isolated from a foetal brain cDNA library. The proteins
CC and coding sequences of the invention can be used in the isolation of
```

```
CC similar genes and proteins, in the elucidation of their function in vivo,
CC and to treat a number of conditions. It is possible that they may have
CC uses as nutritional supplements, as cytokine or cell proliferation
CC factors, in immune modulation, where they may be used to treat immune and
CC autoimmune diseases, as haematopoiesis regulators (treating myeloid or
CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
CC have chemokine or chemotactic activity, haemostatic or thrombolytic
CC activity, or anti-inflammatory activity.
XX
SQ Sequence 1522 AA;

Query Match 5.3%; Score 134.5; DB 21; Length 1522;
Best Local Similarity 21.4%; Pred. No. 0.0038;
Matches 84; Conservative 48; Mismatches 170; Indels 91; Gaps 21;

QY 62 SLSKEA---PREHL-----DHQAAHQPPRRFRQETGHSLSQRDFFRSFLLD-----LPN 109
Db 241 nitreakrppkeefgmngdhtiksq---tprsvhekrvpqeqadaak-fmaqtesgvee 296
QY 110 FPDLSKADIN-GQXPNIQVTIEV-----VDGPDSEADKDOHPENKPSWSVSPDWRAMW 162
Db 297 wsgwstcscvctcgqsgvtrtcvpygthcsqplresvncntalcpvhgv-----weews 352
QY 163 QRSLSLARANSQDDYYXDSTDDSNFLNPPRGWDHTAPGHRFTETKQDPEYDSTDE-G 221
Db 353 pwslcstcgrgtrtrsc-----ppqyggprcpgpethhhkpcnialcpvdgqw 404
QY 222 DWSLWSVCSVTCGNGNQKTRSC-----GYACTA--TESRTCDRPNCPGIEDTFTTAAT 273
Db 405 ewsswqscvtsngtqgrsrqtaahgscrgpwaesrecynpec-----452
QY 274 EVSLLAGSEEFNATKLFVDTSCE-----RWMSCKSEFLK-KYMHKVMNDLPSCP--C 324
Db 453 -----tangqwnqghwscskscdggwerrirtcgavitgqccgsevrrecseqrc 507
QY 325 SYPTEVA---YSTADIFDRKRDFRWK-----DASGPKEK---LEIY-----KPTARYC 368
Db 508 papyeicpedylnsmvwkrtpgdlafnqcpinatgttsrrcslnhgvafwqpsfarc 567
QY 369 IRS-----MLSLESTTLAAQHCCYGDNMQLITR 396
Db 568 lsneyrhlqhsikehlaqgrmlagdgmsqvtk 600

RESULT 13
ABB62815
ID ABB62815 standard; Protein; 763 AA.
XX
AC ABB62815;
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 15237.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
```



Db 1581 cvddn-----knevgardcvakrpyvdresclqpcyvvitge----- 1619  
QY 345 DFRWKDAS-----GPKREKL-----EIV--KPTARYCIRSMLSLESTTLAAQHCCY 387  
Db 1620 ---wsecsvtcgkykqrlvscseiytgkenyeysyqttincpgtqppsvhpcy 1670

RESULT 15  
ABG22373  
ID ABG22373 standard; Protein; 1328 AA.  
XX  
AC ABG22373;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #22364.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS86560.  
XX

New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity -  
XX  
PS Claim 20; SEQ ID NO 52732; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

XX Sequence 1328 AA;  
Query Match 5.2%; Score 130; DB 22; Length 1328;  
Best Local Similarity 21.8%; Pred. No. 0.0083;  
Matches 63; Conservative 26; Mismatches 102; Indels 98; Gaps 13;

QY 192 PPRGWDHTAPGHRFTETKD---QPEYDSTDGE-GDWLSWVSVCTCGNGNKRTRSC--- 244  
Db 543 ppldgtmcapgkchcfkgchciwltptdlikrdgswgawspfgscstcgvgkfrtrqcdnp 602  
QY 245 -----GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVEDTDSC 297  
Db 603 hpanggrtcsqglaydfqlcsrqcdpsladfr-----eeqc 638  
QY 298 ERWMSCKSEFLKYMVKVMDLPSCPCSYPTVEAYSTADIFDIKRKDFRW-----KDAS 352  
Db 639 rqw-----diylfeghdqahhwlphhrda- 662  
QY 353 GPKEKLEIYKPTARYCIRSMLSLESTTLAAQHCCYGDNMQLITRG--KGAGTPNLISTEF 410  
Db 663 --kerchlycesre--tgevsvsmkrmvhdgtcrsykdafslcvrgdcrkvgcdgvg--- 715  
QY 411 SAEHLHYKVDLVPWLIICKGDWSRYNEARPPNNGQKCTESPSDEDIYIKOFQ 459  
Db 716 sskqedkcgv-----cggdnshckvkv-----gtftrspkkgv1kmfe 754

Search completed: April 25, 2002, 14:46:27  
Job time: 9375 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 10:12:56 ; Search time 53.57 seconds  
(without alignments)  
6396.467 Million cell updates/sec

Title: US-09-863-824-1

Perfect score: 1395

Sequence: 1 atgtgtgcctggcgccga.....aagagccagggaattattaa 1395

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCRUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50.6	3.6	1155	2	US-08-387-942C-20
2	50.6	3.6	12588	2	US-08-387-942C-1
3	49.4	3.5	1155	2	US-08-387-942C-7
4	49	3.5	1149	4	US-09-422-487-6
5	48.8	3.5	1155	4	US-09-422-487-4
6	48.8	3.5	1158	3	US-08-844-188-45
7	48	3.4	2220	4	US-08-765-907A-14
8	48	3.4	2888	4	US-08-765-907A-1
9	47.8	3.4	1155	2	US-08-387-942C-18
10	46.8	3.4	1051	4	US-09-245-041-10
11	46.8	3.4	2419	4	US-09-245-041-8
12	46.8	3.4	2625	4	US-09-245-041-18
13	46.8	3.4	4072	4	US-09-245-041-16
14	46.8	3.4	8589	4	US-09-245-041-14
15	46.8	3.4	8827	4	US-09-245-041-1
16	45.8	3.3	2713	3	US-08-804-439A-13
17	45.8	3.3	2713	3	US-08-720-229-13
18	44.6	3.2	1176	2	US-08-387-942C-17
19	44.6	3.2	1635	3	US-09-234-332-4
20	43.8	3.1	477	4	US-09-135-994-1
21	43.4	3.1	3624	1	US-07-951-715A-6
22	43.4	3.1	3624	2	US-08-459-448A-6
23	43.4	3.1	3624	3	US-08-459-595A-6
24	43.4	3.1	3624	3	US-08-459-504B-6
25	43.4	3.1	3624	3	US-08-459-444-6
26	43.4	3.1	3624	3	US-09-053-549-7
27	43.4	3.1	3624	4	US-09-547-422-6

28	42.8	3.1	1465	4	US-09-338-671-1	Sequence 1, Appli
29	42.6	3.1	3472	6	5244792-2	Patent No. 5244792
30	42	3.0	4377	3	US-08-911-853-28	Sequence 28, Appl
31	42	3.0	4377	4	US-09-479-409-28	Sequence 28, Appl
32	42	3.0	4377	4	US-09-479-453-28	Sequence 28, Appl
33	41.8	3.0	8854	3	US-09-053-349-1	Sequence 1, Appli
34	41.6	3.0	4162	2	US-08-459-448A-26	Sequence 26, Appl
35	41.6	3.0	4162	3	US-08-459-595A-26	Sequence 26, Appl
36	41.6	3.0	4162	3	US-08-459-504B-26	Sequence 26, Appl
37	41.6	3.0	4162	3	US-08-459-444-26	Sequence 26, Appl
38	41.6	3.0	4162	4	US-09-547-422-26	Sequence 26, Appl
39	41.6	3.0	4165	1	US-07-951-715A-26	Sequence 26, Appl
40	41.4	3.0	1155	2	US-08-387-942C-21	Sequence 21, Appl
41	40.8	2.9	844	4	US-08-998-416-349	Sequence 349, App
42	40.8	2.9	2004	1	US-08-471-033-18	Sequence 18, Appl
43	40.8	2.9	2004	2	US-08-471-044-18	Sequence 18, Appl
44	40.8	2.9	2004	2	US-08-463-483A-18	Sequence 18, Appl
45	40.8	2.9	2004	2	US-08-471-046A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-08-387-942C-20  
; Sequence 20, Application US/08387942C  
; Patent No. 5939289  
; GENERAL INFORMATION:  
; APPLICANT: ERTESVAG, HELGA  
; APPLICANT: VALLA, SVEIN  
; APPLICANT: SKJAK-BRAEK, GUDMUND  
; APPLICANT: LARSEN, BJORN  
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES  
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/387,942C  
; FILING DATE: 09-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1809-106P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-387-942C-20

Query Match 3.6% Score 50.6; DB 2; Length 1155;  
Best Local Similarity 46.1%; Pred. No. 0.009;  
Matches 170; Conservative 0; Mismatches 199; Indels 0; Gaps 0;  
Qy 896 gctgagtgctgcaaaagcgagttcttaagaagatcatgcacaaagtgatgaatgacc 955

```

,
,
, NAME/KEY: CDS
, LOCATION: 290..1951
,
,
, NAME/KEY: CDS
, LOCATION: 2227..6438
,
,
, NAME/KEY: CDS
, LOCATION: 6702..9695
,
,
, NAME/KEY: CDS
, LOCATION: 9973..12588
,
,
, US-08-387-942C-1

```

	Query Match	3.6%	Score	50.6;	DB 2;	Length	12588;
	Best Local Similarity	46.1%;	Pred.	No. 0.018;			
	Matches	170;	Conservative	0;	Mismatches	199;	Indels 0; Gaps
Qy	896	gcctgatgagctgtcaaaacggagtctcttaaagaagtcatacacaaggctgatgaatgacc	955				
Db	7246	GCGGGGTTCGAGAACACAACTCTCGTACACACGACGCCACGGCTTCAACATCGTCA	7305				
Qy	956	tgccagctgcacctgtccctactccccacttgagttggctcacgacgcccgaatatctcg	1015				
Db	7306	CCAGCACCAACGACTTCTGTTCTTGAGCAACAACGTCGCCTACGGCAAGCGGCGCGGCC	7365				
Qy	1016	accgcataaacgcgaaggacttccgtggaaggacgcacgagcgcccaaggagaagctcg	1075				
Db	7366	TGFTGGTTCGACACGCGCTCGTCGACGTGGCGACCCCCCTACGACATCTGTATCGACGGCG	7425				
Qy	1076	agattctaagagcccatgcccgtactgcctcgcgtccctactgctccctggagagaccac	1135				
Db	7426	GCGCCTACTACGACAAACGGCCTTGGAAGCGGTGCAGATCAAGATGGCCCCACGACGTCAACC	7485				
Qy	1136	cgttygcggcaagcagctgctgtctacggggacacacatgcagctcatcaccaggggccaagg	1195				
Db	7486	TGCAGAAGCGCGAGATTACGGCAACGGCCTATACGGGGTGGCGCTCTACGGCGCCGAGG	7545				
Qy	1196	ggcggcgacgcccacactcatcagcacagtgattcttcgcggagctccactacaagtg	1255				
Db	7546	ATGTGCAGATCCTCGACAACCTACATCCACGACAATTCGAGAACGGTTCTTACGCGGAAA	7605				
Qy	1256	acgtctctgc	1264				
Db	7606	TCTCTCTGC	7614				

```

3
RESULT
US-08-387-942C-7
; Sequence 7, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: BRTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O.BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;

```

```

RESULT 2
US-08-387-942C-1
; Sequence 1, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Azotobacter vinelandii
; STRAIN: E

```



```

; MOLECULE TYPE: DNA (genomic)
; US-08-844-188-45
Query Match 3.5%; Score 48.8; DB 3; Length 1158;
Best Local Similarity 45.5%; Pred. No. 0.023;
Matches 173; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 887 gctgtgagcgtgagtgagtgcaaaagcaggtctttaaagaagtacatgcacaaggtga 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 530 GCTGGACCCCTCGTGGCGTGCATCATGTTCAAGACTTCAAGATCGACAAGAACCCAGA 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 947 tgaatgacctgcccagctgccccctgctctacccccactgagtgccctacacagcgccg 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 590 TCAAGACCAACCCCGTACTATCTTCAAGAAATACAAAGTACTGGAACTCTGGCCAAAGGCT 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1007 acatcttcgaccgcatcaagcgcaaggacttccgctgggaagcagcgccagcgcccaagg 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 650 CCAAGCTGTCCCTCTCCCGCACCAAGCGCAGCTACGACTACGAGTGGGGCACCAGAGA 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1067 agaagctggagatctacaagccccactgcccgggtactgcatcccgctccatgctgctcctgg 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 710 AGAACCAAGAACCAACCATCATCAACACCGTGGGGCTGCAGATCAACATCGACTCGGGGA 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1127 agagcaccagctggcgccagcagcactgctgtctacgcgacacacatgctcatcacca 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 770 TGAAGTTGAGGTGCCGGAGGTGGGGCGGCACCGAGGACATCAAGACCCAGCTCACCG 829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1187 ggggcaaggggggcgccaccccaactcatcagcaccaggttctccgcgagagctccact 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 830 AGGAGCTGAAGTGGAGTACTCCACCGAGACCAAGATCATGACCAAGTACCAGGAGCACT 889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1247 acaagtgagcgtctcctgcc 1266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 890 CCGAGATCGACAACCCGACC 909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-08-765-907A-14
; Sequence 14, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: THIBAUT, Denis
; APPLICANT: BLANC, Veronique
; APPLICANT: BAMA-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; TITLE OF INVENTION: Mutasynthesis
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2220
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
; US-08-765-907A-14

Query Match 3.4%; Score 48; DB 4; Length 2220;
Best Local Similarity 46.9%; Pred. No. 0.042;
Matches 150; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 950 atgacctgccagctgccccctgctcctaccactgagtgccctacagcgccgagcaga 1009
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1658 aggagcgccgctcaaggcctcctcgcgcggcgaggaagaccgagcgagaacctga 1717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT

US-08-387-942C-18  
; Sequence 18, Application US/08387942C

RESOLUTION 8  
US-08-765-907A-1  
: Sequence 1, Application US/08765907A

APPLICANT: BLANC Veronique

APPLICANT: THIBAUT, Denis

; APPLICANT: **BLANCHE, Francis**

APPLICANT: BARRIERE, Jean-Claude

APPLICANT: FAMECHON, Alain

APPLICANT: PARIS, Jean-Marc

; TITLE OF INVENTION: Streptogramins

; FILE REFERENCE: Streptogramin genes

CURRENT FILING DATE: 1997-03-20

```

; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patatix Ver 3.0

```

; SEQ ID NO 1

; TYPE: DNA

US-08-765-907A-1

Query Match	3.4%	Score 47.8;	DB 2;	Length 1155
Best Local Similarity	44.8%	Pred. No. 0.039;		

Matches 184; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

1000  
 950  
 900  
 850  
 800  
 750  
 700  
 650  
 600  
 550  
 500  
 450  
 400  
 350  
 300  
 250  
 200  
 150  
 100  
 50  
 Db

0v 1010 tcttcgaccgcacatcaagcgcgaaggacttcccgctggaaagcacgcccccagaagana 1069

1000

Db 185 TgATCgtcgacCTggtccgcaagacCTcgccaggtctgcacatcggctccgtccacg 24A

244

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

Qy 1070 agctggagatctacaagcccaactgcccggtagctgcatccgctccatgctgtccctggaga 1129

Db 245 taccgggcctgttcgagggtggagacctacgccaccgtccaccagctcgtcacacacgtcc 304

-----

0v 1130 acaccacgcctggccggcacacacactgctactacgggcacacacatgcacactcaccaccagg 1189

1185

Db 305 gcggccgctggcggccgagctctccgcgcccgccggtacgggcccgttccccggcg 364

Qy 1190 gcaagggggcgggcacgcccacctcatcagcacccgagttctccgcgagctccactaca 1249

1

\_\_\_\_\_









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2002, 12:13:27 ; Search time 28.09 Seconds  
(without alignments)  
403.470 Million cell updates/sec

Title: US-09-863-824-2  
Perfect score: 2517  
Sequence: 1 MVRLAELLGGLLLTLH.....CTSPSDEYIKQFQEARAY 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA:\*
- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/iaa/6CTUS\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	5.1	239	5	PCT-US93-01652-1
2	127	5.0	1170	1	US-08-313-288B-20
3	117.5	4.7	807	1	US-07-862-021B-10
4	117.5	4.7	807	1	US-08-313-288B-10
5	117.5	4.7	807	5	PCT-US93-03164-10
6	115.5	4.6	1156	3	US-08-996-083-1
7	115.5	4.6	1156	4	US-09-429-516-1
8	115.5	4.6	1156	4	US-09-429-516-3
9	115	4.6	788	2	US-08-918-914-4
10	115	4.6	1172	1	US-08-313-288B-19
11	113	4.5	218	3	US-08-985-526-1
12	113	4.5	441	3	US-08-985-526-3
13	111	4.4	469	1	US-08-313-288B-15
14	110.5	4.4	1785	4	US-09-341-587-3
15	106.5	4.2	568	1	US-07-862-021B-14
16	106.5	4.2	568	5	PCT-US93-03164-14
17	106	4.2	898	2	US-08-808-982-5
18	106	4.2	898	4	US-09-306-902A-5
19	105	4.2	802	1	US-07-862-021B-12
20	105	4.2	802	1	US-08-313-288B-12
21	105	4.2	802	5	PCT-US93-03164-12
22	104.5	4.2	559	1	US-08-313-288B-14
23	103	4.1	943	2	US-08-808-982-7
24	103	4.1	943	4	US-09-306-902A-7
25	101	4.0	60	1	US-07-646-531D-12
26	101	4.0	60	2	US-08-488-273-12
27	101	4.0	60	6	5426100-12

Patent No. 5426100

ALIGNMENTS

RESULT 1  
PCT-US93-01652-1  
; Sequence 1, Application PC/TUS9301652  
; GENERAL INFORMATION:  
; APPLICANT: Bouck, Noel P.  
; APPLICANT: Polverini, Peter J.  
; APPLICANT: Good, Deborah J.  
; TITLE OF INVENTION: Method and Composition for  
; TITLE OF INVENTION: Inhibiting Angiogenesis  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut  
; STREET: 100 South Wacker Drive, Suite 960  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606-4002  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/01652  
; FILING DATE: 19930222  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/841.656  
; FILING DATE: 24-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/464.369  
; FILING DATE: 12-JAN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pentress, Susan B.  
; REGISTRATION NUMBER: 31,327  
; REFERENCE/DOCKET NUMBER: 92005-PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312)-456-8000  
; TELEFAX: (312)-456-7776  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 239 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
PCT-US93-01652-1

Sequence 8, Appli  
Sequence 7, Appli  
Sequence 3, Appli  
Sequence 7, Appli  
Sequence 7, Appli  
Sequence 13, Appli  
Patent No. 5426100  
Sequence 97, Appli  
Sequence 97, Appli  
Sequence 97, Appli  
Sequence 97, Appli  
Sequence 97, Appli  
Sequence 97, Appli  
Sequence 103, App  
Sequence 5, Appli  
Sequence 4, Appli



CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,288B  
FILING DATE: January 5, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28 678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 807 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-313-288B-10

Query Match 4.7%; Score 117.5; DB 1; Length 807;  
Best Local Similarity 21.0%; Pred. No. 0.0035;  
Matches 82; Conservative 35; Mismatches 137; Indels 137; Gaps 17;

Qy	83	RPRFRQRTGHPISLQRDFPFRSLDLLENFPD-----LSKADINGOXPNI---QVT	128
Db	368	KPTIQEKIRPLTSLHQPSPFYD----PEGGSITQVARVVIERTARKGEQCNIVPDNVD	423
Qy	129	IEYVDGPDSEADKDHOPE-----NKPSW-----SVSPSD- 157	
Db	424	DIVADLAPEEKDEDDTPETCIYNWSFWACSSTCEKKRMRQLKAQLDLSPVCPT	483
Qy	158	-----WRAWORSLARANSODDYXYDSTDSDSNFLNPPR	194
Db	484	QDFQPCMGPCSDGDGSTCTMSEMTWSCVSVCGMRSERYVKQFPEDCSVMCLP--	541
Qy	195	GWDHTAPGHFTFKDOPEVDSTDGBG----DWLSMVCSVTCGNQNRTR-----SC	244
Db	542	-----TEETKTVNEECSPSSCLVTEWGEDDCSATCGMGMKRRHRWKMSPAD	591
Qy	245	GYACTA--TESR7CDPNCPGIEDTFRTAAATEVSLLAGSEERNATKLFVEDTDTCERWNS	302
Db	592	GSCKAETSQAEEKMMPECHTIP-----CLLSPWSEWSDCSV-----TCCKGMR	635
Qy	303	CKSEFLX-----KYMHKVMDNLPSCP--CSYPTEVAYSTAD-----IFPR	340
Db	636	TRQRLMSLAELGCNDLELQAEKCM-LPECPIDELSELSEWSOWSECNKGCGHMIRTR	693
Qy	341	IKRKDFRWKDAAGPKLEYTKPTARYCIRS	371
Db	694	TIQMEPOFGGAPCP-ETVQRKKCRARKCLR	723

```

RESULT      5
PCF-US93-03164-10
: Sequence 10, Application PC/TUS9303164
: GENERAL INFORMATION:
: APPLICANT: Jessell, Thomas M
: APPLICANT: Klar, Avihu
: TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A

```

```

; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
;
; NUMBER OF SEQUENCES: 20
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Cooper & Dunham
;
; STREET: 30 Rockefeller Plaza
;
; CITY: New York
;
; STATE: New York
;
; COUNTRY: USA
;
; ZIP: 10112
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatenLin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: PCT/US93/03164
;
; FILING DATE: 19930402
;
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER:
;
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: White, John P
;
; REGISTRATION NUMBER: 28,678
;
; REFERENCE/DOCKET NUMBER: 40028
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (212) 977-9550
;
; TELEFAX: (212) 664-0525
;
; TELEX: 422523 COOP UI
;
; INFORMATION FOR SEQ ID NO: 10:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 807 amino acids
;
; TYPE: AMINO ACID
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; PCT-US93-03164-10

```

Query Match	4.7%;	Score 117.5;	DB 5;	Length 807;
Best Local Similarity	21.0%;	Pred. No. 0.0035;		

[illegible]

RESULT 6  
US-08-996-083-1



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,083
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0420 US
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1156 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SATPF1002
; CLONE: 422069
; US-09-429-516-3

```

```

Query Match 4.6%; Score 115.5; DB 4; Length 1156;
Best Local Similarity 25.8%; Pred. No. 0.01;
Matches 41; Conservative 13; Mismatches 50; Indels 55; Gaps 9;

QY 154 PSP-----DWRAWQ-----RSLSLARANSQ-----DQDYXY 180
Db 46 PPALEDWEAESEWTFNVDFHFGDGFESLAAIRFYGPVPCRPRLAERTTDMAL 105

QY 181 DSTSDSNFLNPRG-W--DHTAPG-----HRTFTKQDQPEVSDTDGSDSLWSVCS 230
Db 106 PSVAGERVHLNPRGFWCLNRPGRRCNSYHVRCFPLEASW-----GANGPWGPCS 159

QY 231 VTCGNGNKRTRSC---GYACTA---TESRTCDRPNCPG 263
Db 160 GSCGPGRRRLRRHCPSPAGDACPRLPQAQKCVRCPCPG 198

```

```

RESULT 9
US-08-918-914-4
; Sequence 4, Application US/08918914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

```

```

; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1070094
; US-08-918-914-4

Query Match 4.6%; Score 115; DB 2; Length 788;
Best Local Similarity 19.1%; Pred. No. 0.0061;
Matches 62; Conservative 36; Mismatches 120; Indels 106; Gaps 11;

QY 65 KEAPREHLHQA-----AHQFPFRPRQETGHPSLQDRFPRSELL 105
Db 230 QOREQQRHREHQAQLQHQQQQQQQQQQQQPPQPQPQPQPQPSFSGTHEL 289

QY 106 DLPNFPD-----LSKADINGOXPNQVT-----IEVVDG--PDSEADKDDHPE 146
Db 290 HLQRQRQEQQQQQQQQQQQQQQQQQQQQTTFQCSQIQLOSQVPPQOHHQOQPPQ 349

QY 147 NKPSWSVSPDWRA--WMQSLSLARANSQDQDYDYDSTSDSNFLNPRGWDHTAPGHRT 205
Db 350 QOPELERSPLDQHAQLQQRMSQYREN-----FNORHPARPKADPCPGFC 395

QY 206 FETKQDQPEYD-----STDGEG----- 221
Db 396 APVPAQQRERPTPPVLPVINTATQPLPQPYRPPAPPPACDGGCVNPPVWS 455

QY 222 ----DWSLWSVCSVTCGNGNKRTRSCG-YACTAT--ESRTCDRPNCPGIEDTFTTAATE 274
Db 456 GVMHDSWSTCSTCGDGAKSRRRECSTNNOGADYETPCNLGPGCOTWSEWCENWSTCS 515

QY 275 VSLIAGSEEFNATKLFVETDSDCE 298
Db 516 AS--CGSGQRERTFCHLGNRCE 537

```

```

RESULT 10
US-08-313-288B-19
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-313-288B-19

Query Match 4.6%; Score 115; DB 1; Length 1172;  
Best Local Similarity 22.2%; Pred. No. 0.012;  
Matches 40; Conservative 14; Mismatches 54; Indels 72; Gaps 8;  
QY 146 ENKPSWSPDRAWQSRSLARANGDODYYXDSTSDSNFLNP----- 192  
DB 379 DGEEGWS-----PWAETQCSVT---CGSGTQGRGRSCDVTSNCLGFSIQTRACSLSKCD 431  
QY 193 -----PRGWDHTAP-----GHRFTFETKDOPE 213  
DB 432 THIRODGGWHSWSSCVTCGVGNITRILCNSPVPMGKNCKGSGRETKACQGAP- 490  
QY 214 YDSTGGE-GDWSLWSVCVTCGNGNOKRTRSC-----GYACT--ATESRTCDRPNCP 262  
DB 491 -CPIDGRWSPWSPSACTVTCAGGIRERTVCNSPEPQYGGKACVGDVQERQMCNKRSCP 549

RESULT 11  
US-08-985-526-1  
; Sequence 1, Application US/08985526  
; Patent No. 6080728  
; GENERAL INFORMATION:  
; APPLICANT: Mixson, James A  
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA  
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
; TITLE OF INVENTION: THERAPY  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Connolly, Bove, Lodge, & Hutzel  
; STREET: 1220 Market Street, P.O. Box 2207  
; CITY: Wilmington  
; STATE: Delaware  
; COUNTRY: U.S.A.  
; ZIP: 19899  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/985,526  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,845  
; FILING DATE: 16-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMorrow Jr., Robert G  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (302) 658-9141  
; TELEFAX: (302) 658-5613  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 218 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

US-08-985-526-1

Query Match 4.5%; Score 113; DB 3; Length 218;  
Best Local Similarity 33.8%; Pred. No. 0.0013;  
Matches 25; Conservative 8; Mismatches 21; Indels 20; Gaps 3;  
QY 193 PRGWDHTAPGHRFTFETKDOPEYDSTGEGDWSLWSVCVTCGNGNOKRTRSC---GYAC- 248  
DB 70 PRCW-----PDSADDDGWSWPSEWTSCTSTSCGNGIQGRGRSCDSLNNRCE 114  
QY 249 -TATESRTCDRPNCP 261  
DB 115 GSSVQTRTCHIQEC 128

RESULT 12  
US-08-985-526-3  
; Sequence 3, Application US/08985526  
; Patent No. 6080728  
; GENERAL INFORMATION:  
; APPLICANT: Mixson, James A  
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA  
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
; TITLE OF INVENTION: THERAPY  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Connolly, Bove, Lodge, & Hutzel  
; STREET: 1220 Market Street, P.O. Box 2207  
; CITY: Wilmington  
; STATE: Delaware  
; COUNTRY: U.S.A.  
; ZIP: 19899  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/985,526  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,845  
; FILING DATE: 16-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMorrow Jr., Robert G  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (302) 658-9141  
; TELEFAX: (302) 658-5613  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 441 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-985-526-3

Query Match 4.5%; Score 113; DB 3; Length 441;  
Best Local Similarity 33.8%; Pred. No. 0.0039;  
Matches 25; Conservative 8; Mismatches 21; Indels 20; Gaps 3;  
QY 193 PRGWDHTAPGHRFTFETKDOPEYDSTGEGDWSLWSVCVTCGNGNOKRTRSC---GYAC- 248  
DB 293 PRCW-----PDSADDDGWSWPSEWTSCTSTSCGNGIQGRGRSCDSLNNRCE 337  
QY 249 -TATESRTCDRPNCP 261  
DB 338 GSSVQTRTCHIQEC 351

RESULT 13  
US-08-313-288B-15

```

Query Match      4.4%; Score 111; DB 1; Length 469;
Best Local Similarity 26.1%; Pred. No. 0.007;
Matches 36; Conservative 11; Mismatches 51; Indels 40; Gaps 7;

QY 145 PENKPSVSPDWRWQRSLSLARANGSQDDXYDSTSDSDSFLNPPR--GWD----- 197
Db 74 PCRSPRWSL-----WSTWAPCSV-----TCSEGSOLRYRRCVGVWNGQCSG 113
QY 198 HTAPGHRTFE---TKDQPEYDSTDGEGDWLSWSCSVTCGNGNOKRTRSCGYAC----- 248
Db 114 KVAPGTLEWLOACEDQOCCPEMGSGWGWPEFCSVTCSKGTTRRRACNHPAPKCGGH 173
QY 249 ---TATESRTCDRPN-CP 262
Db 174 CPGQAQSEACDTQOVCP 191

RESULT 14
US-09-341-587-3
; Sequence 3, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: PCT/DE98/00096
; EARLIER FILING DATE: 1998-01-09
; NUMBER OF SEQ ID NOS: 12

```

```

RESULT 15
US-07-862-021B-14
; Sequence 14, Application US/07862021B
; Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 568 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-021B-14

```

Query Match 4.28; Score 106.5; DB 1; Length 568;

Best Local Similarity 21.3%; Pred. No. 0.028;			
Matches	70; Conservative	28; Mismatches	122; Indels 109; Gaps 14;
Qy	93	PSLQDFPRFL-LDLNFPDLKADINGXPNIQVTIE-----VV	132
Db	129	PTIPQDKIRPLTSLDHPQSPMTRG--GPIIPIARVVIERARKGEOCNIPDNVDDIVA	186
Qy	133	DGPDSEADKQHP-----NKPSW-----SVPSPD-----	157
Db	187	DLVTEKDEDDTPTCYISNWPMSACSSATCDKGRMRQMLKAQLDLSVPCCPDQDFE	246
Qy	158	-----WRAWQRLSLARANSQDQYXDSTSDSNFLNPPRGWDH	198
Db	247	PCMGPGCSDDDEASTCMKSEWITNSPCASCGMGIEVREYVVKQFPEDGSLCKVP-----	300
Qy	199	TAPGHRFTETKQPEYDSTDEG-----DWLSVSVCTCGNGNOKR-----TRSCGYAC	248
Db	301	-----TEETEKICIVNEECEPSSCIVTEWAEWECSATCRMGKMKRHRMIKMTPADGSMC	354
Qy	249	TA--TESITCDRPNCPGIEDTFR--TAATEVSLLAGSEEFNATKLFVTDSCERWMSCK	304
Db	355	KADTVEVEKCMMPCHTIPCVLSPWSEWSDCSVTCGKGTTRQRMLKSPSELGD-----CN	410
Qy	305	SEFLKKYMHKVMNDLPSCPSYP-TEVAY	332
Db	411	EELELKQVEKCM--LPECPISELTEWSY	437

Search completed: April 25, 2002, 14:47:23  
Job time: 9236 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 10:06:26 ; Search time 1542.46 Seconds  
(without alignments)  
12206.639 Million cell updates/sec

Title: US-09-863-824-1

Perfect score: 1395

Sequence: 1 atgtgtgcctggcgccga.....aagagccagggaattattaa 1395

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_gss:\*
  - 13: em\_gss\_hum:\*
  - 14: em\_gss\_inv:\*
  - 15: em\_gss\_pln:\*
  - 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	577	41.4	593	9 AW873508	AW873508 ho60h09.x
2	539.2	38.7	644	10 BF472502	BE472502 UI-M-BH3-
3	413.4	29.6	544	10 BE947736	BE947736 UI-M-BH3-
4	407	29.2	691	10 BG162342	BG162342 de33e08.y
5	399	28.6	968	10 BF302850	BF302850 602032745
6	384.2	27.5	425	9 AA884248	AA884248 am32b09.s
7	365	26.2	645	10 BJ060213	BJ060213 BJ060213
8	345.2	24.7	356	10 W88257	W88257 zd91f09.r1
9	326.8	23.4	383	9 AW426875	AW426875 61687 MAR
10	284	20.4	379	9 AW495535	AW495535 UI-M-BH3-
11	278.2	19.9	549	10 BI535390	BI535390 398882 MA
12	260.2	18.7	479	9 AW435642	AW435642 74577 MAR
13	259.8	18.6	682	10 BF571346	BF571346 602077485
14	244.4	17.5	446	10 BM191035	BM191035 ga331b08.
15	242	17.3	666	10 BI891452	BI891452 2F637-3-0
16	237.8	17.0	715	10 BI704199	BI704199 zfshtARAM
17	234.6	16.8	266	9 AW435639	AW435639 74571 MAR

18	233.4	16.7	1032	9 AL545675	AL545675
19	211.2	15.1	643	9 BB322233	BB322233
20	159.6	11.4	673	12 AO606601	AO606601 HS-2095.B
21	157.8	11.3	275	10 WL4682	WL4682 mb30h04.r1
22	149.8	10.7	827	10 BG739780	BG739780 602630502
23	147	10.5	914	12 CNS021K6	AL199023 Tetradon
24	142.2	10.2	311	9 BB586477	BB586477 BB586477
25	137	9.8	679	9 AL659826	AL659826 AL659826
26	135.6	9.7	463	10 BG037854	BG037854 dc69a05.y
27	129.4	9.3	141	9 AW435643	AW435643 74579 MAR
28	116.8	8.4	608	10 BJ071721	BJ071721 BJ071721
29	116.4	8.3	133	9 AA435648	AA435648 74587 MAR
30	93	6.7	747	9 AA447294	AA447294 AC.E04R P
31	89.2	6.4	298	9 AI430783	AI430783 mh66a04.x
32	81.8	5.9	693	10 BE287313	BE287313 601093984
33	81.4	5.8	546	10 BI442837	BI442837 da191a08.
34	81	5.8	589	12 AZ995663	AZ995663 2M0281N06
35	80.4	5.8	353	9 AI070670	AI070670 UI-R-C2-m
36	80.2	5.7	702	10 BJ082998	BJ082998 BJ082998
37	72	5.2	282	10 R19947	R19947 yg30h05.r1
38	64.2	4.6	433	10 BI449830	BI449830 dag32a03.
39	51.4	3.7	639	10 BG344947	BG344947 HVSME001
40	50.6	3.6	416	10 BE352595	BE352595 WHE0425_E
41	50.6	3.6	595	9 AL504377	AL504377 AL504377
42	50.6	3.6	677	10 BE415842	BE415842 MUG001.F0
43	50.6	3.6	771	12 AO861404	AO861404 nbeb0016H
44	50	3.6	474	10 BM347288	BM347288 MEST275-D
45	50	3.6	506	9 AI691386	AI691386 606015B10

## ALIGNMENTS

RESULT 1	AW873508	583 bp	mrna	linear	EST 22-MAY-2000
LOCUS	ho60h09.x1	Soares_NFL_T_GBC_S1	Homo sapiens	cDNA clone	
DEFINITION	IMAGE:3041825 3' similar to TR:095432 095432 HYPOTHETICAL 72.5 KD PROTEIN. ;, mRNA sequence.				
ACCESSION	AW873508				
VERSION	AW873508.1	GI:8007561			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 583)				
TITLE	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Possible reversed clone: similarity on wrong strand Seq primer: -40UP from Gibco High quality sequence stop: 463. Location/Qualifiers 1. 583 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3041825" /clone_lid="Soares_NFL_T_GBC_S1" /lab_host="DH108" /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung Nhlh19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made				

## FEATURES

source	1. 583
Location/Qualifiers	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3041825" /clone_lid="Soares_NFL_T_GBC_S1" /lab_host="DH108" /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung Nhlh19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made



Db	404	ATTGCCACAGAAATCCAGGACCTGTGACCGTCCAAACTGCCAGGAATTAAGATACATTTC	345
Qy	805	aggacagctgccaccggaagtgagctctcttgcgggaagcagagaggttaatgcaccacaaa	864
Db	344	AGGACAGCGGCCACTGAAGTGAGTCTGCTTGCGGGAAGTGAGGAGTTTAATGCCACCAAG	285
Qy	865	ctgtttgaattgacacagacagctgtgagcgtctggatgaagctcaaaagcgaattctta	924
Db	284	TTGTTTGAAGTTGACATGGACAGCTGTGAGCGATGGATGAGCTGCAAAAGTGAGTTCTTA	225
Qy	925	aagaagtacatgcacaaggtgatgaatgacctgccagctgcccctgcctctaccccact	984
Db	224	AAGAAATACATGCACAAGGTGATCAACGACCTGCGCCAGCTGCCCCCTGCTCTCACTCCCTACT	165
Qy	985	gaggtggcctacagcacgcccagacattcttcagccgacatcaagcgcgaaggacttcgcgtgg	1044
Db	164	GAGGTGGCCCTACAGCAGCTGAGCATCTTTGACCGCATCAAGCGCAAGGACTTCCGATGG	105
Qy	1045	aaggacccacgaggcccaaggaagctggagatctcacagcccactgcccgtactgc	1104
Db	104	AAGGATGCTAGTGGGCCCAAGAGAAACTAGAGATCTACAGCCTACTGCTCGGTACTGC	45
Qy	1105	atccgctccatgctgtctccctggagagcaccacacgctggcgccac	1147
Db	44	ATCCGCTCTATGCTGTGCTCCCTGGAGAGTACCACACTGGCCCGCTC	2

### RESULT 3

BE947736	544 bp	linear	EST 03-OCT-2000
LOCUS	BE947736	UT-M-BH3-axd-h-03-0-UI.s1	musculus cDNA clone
DEFINITION	UT-M-BH3-axd-h-03-0-UI 3', mRNA sequence.		
ACCESSION	BE947736		
VERSION	BE947736.1	GI:10525495	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 544)		
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B.		
JOURNAL	Normalization and subtraction: two approaches to facilitate gene		
MEDLINE	discovery		
COMMENT	Genome Res. 6 (9), 791-806 (1996)		
	97044477		
	Contact: Chin, H		
	National Institute of Mental Health		
	6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD		
	20892-9643, USA		
	Tel: 301 443 1706		
	Fax: 301 443 9890		
	Email: meste@mail.nih.gov		

**FEATURES**  
**source**

[illegible]

10

[illegible]









JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 20  
and -minmatch 12 options.

## PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCACGCTACGACG

Plate: 24 row: L column: 20

Seq primer: ATTAGTGTGACACTATAG.

FEATURES  
source

1..682  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4252124"  
/clone\_lib="NIH\_MGC\_62"  
/tissue\_type="melanotic melanoma, high MDR"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGCGCGGCGGACATG-dT(30)BN-3',  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 171 a 183 c 212 g 116 t  
ORIGIN

Query Match 18.7%; Score 260.2; DB 9; Length 479;

Best Local Similarity 93.4%; Pred. No. 5e-42;

Matches 268; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 359 ggcagrvtccaaatccaggctaccatagagtggtcgacggtcctgactctgaagcag 418

Db 12 GGCAGATCCAAATATCCAGGTCACCATAGAGTGGTGGACGGTCTCAATGTGAAGCAG 71

QY 419 ataaagatcagatcccgagagaataagccagctgttcagtcctccatcccgactggcggg 478

Db 72 ATAAAGATCAGATCCCGAGAGTAAGCCAGCTGGTCACTCCCATCCCTGACTGGCGGG 131

QY 479 cctggtggcagagtgctccttggccaggggcaaacagcggggagcagactacragt 538

Db 132 CCTGGTGGCAGAGTCTCTTCTCTGGCAAGAGCCACAGCGGGGACGAGACTACAAGT 191

QY 539 acgacgtactcagacagcagcaactctcaacccccccagggggtggaccatacag 598

Db 192 ACGACGTACTCAGATGACAGCAACTTCTCAACCCCCCGGGGGCTGGGACCATCCGG 251

QY 599 cccagggccacgggactttgaaaccaaagatcagcagaaatgat 645

Db 252 CCCAGGCCACCGGACTTTCGAAACCAAGAGAGCAGCAGCAATATGGT 298

## RESULT 13

BF571346

LOCUS 602077485F1 NIH\_MGC\_62 Homo sapiens cDNA clone IMAGE:4252124 5',

DEFINITION mRNA sequence.

ACCESSION BF571346

VERSION BF571346.1 GI:11645058

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 682)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/DTP

CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1076 row: i column: 21  
High quality sequence stop: 379.

FEATURES  
source

1..682  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4252124"  
/clone\_lib="NIH\_MGC\_62"  
/tissue\_type="melanotic melanoma, high MDR"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGCGCGGCGGACATG-dT(30)BN-3',  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 171 a 183 c 212 g 116 t  
ORIGIN

Query Match 18.6%; Score 259.8; DB 10; Length 682;

Best Local Similarity 88.8%; Pred. No. 6.4e-42;

Matches 420; Conservative 0; Mismatches 32; Indels 21; Gaps 12;

QY 939 caaggtgatgaatgacgtcccgactgccccctgctctaccctactgagtggtgctacag 998

Db 1 CAAGGTGATGAATGACGTGCCAGTGCCTCTTACCCCACTGAGTGGCTACAG 60

QY 999 cagcgccgacatcttcgacgcgcatcaagcgcaaggaactccgtggaggagcgcagcg 1058

Db 61 CACGG-CGACATCTTCGACCGCATCAAGCGCAAGGACTTCGCGTGGGAAGGACGCAGCGG 119

QY 1059 gccca-agggagaagctggagatctacaagccactgccgtactgcctccatgc 1117

Db 120 GCCCATAGGAGAAGCTGGAGATCTACAAGCCCACTGCCGGTACTGTCATCG-TCCATGC 178

QY 1118 tgtccctggagagcaccacgcctggcgccagcactgctgtaaggacacatgcagc 1177

Db 179 TGTCCCTGGAGAGCACCACCGCTGGCGGCACAGAACTGTTG--TACGGGACAAATGCAGC 236

QY 1178 tcataccaggggcaagggggcgccagcccaacctcatcagcaccaggtctcccg 1237

Db 237 TCATCACCGGGGCAAGGGCGGGGAAGC-CCAAACTCATCAGCACCGAGTTCTCCCGCGG 295

QY 1238 agctccact--acaaggtggagctcctgccc-tggattatctg---caaggggtgactgga 1291

Db 296 AGCTCCAATTACAAGGTGGAGCTCTCCCGTGGATTATCTGCAAAAGGTGTGACTGGA 355

QY 1292 g-caggtataacgagggcccgctcccaacaacg-----gacagagtgccagagagc 1344

Db 356 GCCAGGTATACGAGGCGCGGCTCCACAAAGCGGGAACAGACGTGGGCCAGAGAGC 415

QY 1345 ccctcggac-gagacatcatcaagc-agttccaaagggccaggaattataa 1395

Db 416 CCCTCGGACGGAGACTACATCAAGCAAGTTTCCAAGAGGCCAGGACATTTTAA 468

## RESULT 14

BM191035

LOCUS

DEFINITION

5' similar to TR:095432 095432 HYPOTHETICAL 72.5 KD PROTEIN. ;

EST 11-DEC-2001

IMAGE:5077959

```

mRNA sequence.
BM191035          1  GI:17526998
EST.
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis
African clawed frog.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 446)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other_ESTs: daj31b08.xl
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov
Seq primer: -40Rp from Gibco
High quality sequence stop: 442.
FEATURES
             Location/Qualifiers
             1..446
             /organism="Xenopus laevis"
             /db_xref="taxon:8335"
             /clone="IMAGE:5077959"
             /clone_lib="NICHD XGC 001"
             /tissue_type="oocytes"
             /lab_host="DH10B (phage-resistant)"
             /note="vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI;
             Cloned unidirectionally. Primer: Oligo dT. Average insert
             size 2.2 kb. Constructed by Life Technologies."
BASE COUNT   132 a   100 c   116 g   98 t
ORIGIN
Query Match      17.5%; Score 244.4; DB 10; Length 446;
Best Local Similarity 71.4%; Pred. No. 6.9e-39;
Matches 317; Conservative 3; Mismatches 124; Indels 0; Gaps 0;
QY  280  tcattgaagaagattcccccagatccttctccttgatctaccacaaacttccagattc 339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2  TCACTCCACGAGATGGGACGGATCCTTCTCTGGATCTTCCCAATTTTCCCTGACCTC 61
QY  340  tccaaagctgatcatatggcagrtwccaatatccagggtccaccatagaggtggtcag 399
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    62  TCCAAAGTTGATATTAATGCCAAATCCTAATATTCAGTCACCTATTGAAGTGGTTGAT 121
QY  400  ggctcgtactgaacagataaagatcagatccgagataaagccagctgggtcagtc 459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    122  GGACCTGACTCGGACGGGAATGAATGCAAAAGGAGAACATGCCAGCTGGCCGGTT 181
QY  460  ccataccccagctggcggcgtggcagaggtccctgtccttggccaggaggcaaacagc 519
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    182  CCATCACCTGACTGGAGAACTGCTGGCAGAGTTCAGCAACTTTGCCCGTATGAATTAC 241
QY  520  ggggaccagactacragtacagatcactcagacagacagaaacttcccaaccccc 579
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    242  GGGGACCAAGATTACAAATACGACAGCTACTGAGAGCAGCAACTTCCTTAACCCCTCG 301
QY  580  aggggggggaccatacagccccagccagccagcttttgaacacaaagatcagccagaa 639
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    302  GGAGGAAGGAATACACAAGTGCCGCCAGCAACATTTGATACAAAGGAACAACTGAG 361
QY  640  tatgattccacagatggcagggtgagctgagctgctgtctgtctgtctgagcgtcaactgc 699
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    362  TATGACTATGTAGATGGTGGAGAGATTGGAGCAGTTGTGAGTGTGCTGCTGCTGCTG 421
QY  700  ggggaacyggaacgaagacgacc 723

```

```

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    422  GGTAGCGGTAACCAAAAGCGCAC 445

RESULT 15
BI891452
LOCUS
DEFINITION
Z637-3-000707 zebrafish shield stage whole embryo cDNA library
MPMGp637_Danio rerio cDNA clone MPMGP637_15P5:MPMGp637P0515 5',
mRNA sequence.
BI891452          666 bp      mRNA      linear      EST 12-OCT-2001
Z637-3-000707 zebrafish shield stage whole embryo cDNA library
MPMGp637_Danio rerio cDNA clone MPMGP637_15P5:MPMGp637P0515 5',
mRNA sequence.
BI891452          1  GI:16098723
VERSION
KEYWORDS
SOURCE
ORGANISM
zebrafish.
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 666)
Clark, M., Aanstad, P., Hennig, S., Johnson, S.L. and Lehrach, H.
EST sequencing of a zebrafish shield stage cDNA library normalised
by oligonucleotide fingerprinting
Unpublished (2001)
JOURNAL
COMMENT
Contact: Hennig S
Laboraty 123, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
5' EST sequencing of clones from a zebrafish shield stage library,
normalised from 55,000 starting clones by oligonucleotide
fingerprinting
High quality sequence stop: 666.
FEATURES
             Location/Qualifiers
             1..666
             /organism="Danio rerio"
             /db_xref="taxon:7955"
             /clone="MPMGp637_15P5:MPMGp637P0515"
             /clone_lib="zebrafish shield stage whole embryo cDNA
             library MPMGP637"
             /tissue_type="whole embryo"
             /dev_stage="shield stage, 6 hrs post-fertilisation"
             /lab_host="E.coli, XL1 blue MRF"
             /note="Vector: pSport1; Site.1: NotI; Site.2: SalI;
             oligo-dT-NotI primed, SalI adaptors, directionally cloned,
             library normalised by oligonucleotide fingerprinting"
BASE COUNT   186 a   158 c   193 g   127 t
ORIGIN
Query Match      17.3%; Score 242; DB 10; Length 666;
Best Local Similarity 64.4%; Pred. No. 2.2e-38;
Matches 433; Conservative 2; Mismatches 209; Indels 28; Gaps 4;
QY  350  atataatggcagrtwccaatatccaggtccaccatagaggtggtgcaggtcctga-- 407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2  AGATTATATGGCGACAGACCCCATATACCAAGTACCATAGAGTGGTGGACAGTCTGGAAG 61
QY  408  -ctctgaagcagataaagatcagcatccggaggaataagccagctggtcagtcctcc 466
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    62  GTTCTGAGCCAGAGAAAGGAATGGTAAAGAAACAAAGCCCTGGCTGGGCAG-----CTC 115
QY  467  ccgactggcggcctggtgagagaggtccctgtccttgccaggagggcaaacag----- 518
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    116  CTAACGTGAGGAACATGGTGGCAGCGTTCATCATCTCTCTCATCTCTCTGTGTGCTCACTC 175
QY  519  -----cgggaccaggtactacragtacagctacacagtcacagcagacgaacttcc 568
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    176  CGAAGGGGCTGAGGAACAGGATTACCCCTACGAGAGCAACACAGAGGACAGCAACTTCC 235
QY  569  tcaacccccccagggggtgggaccatacagccccagccagcagcttttgaacacaaag 628
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Db 236 TCAAGCCACTCGGAGACTGGGAGAGAGAGTGAAGAGTGAAGCTGGAGCGGGAGACAGAA 295
Qy 629 atcagccagaatatgattccacagatggcgaggtgactgagtgctctctggtctgtctgca 688
Db 296 CCCAGACTGATACGACTACATACACGGGGAAGCGGACTGGAGTGTGGTCTCCATGCA 355
Qy 689 gcgtacctgcgggaacggaacacagaaacggaaccccggtcttctggtctgctgcaactg 748
Db 356 GTGTATCTGTGGAAATGGTAACCAAGAGCGGACAGATCGTGGGTTATGCCCTGCACTG 415
Qy 749 caacagaatcgagagacctgtgaccgtccaaactgcccaggaaattgaagacaccttttagga 808
Db 416 CCACAGAGTCAACGACATGTACATGCCAGCTGTCCAGGGATTGAAGATGCATTCAAGA 475
Qy 809 cagctgccaccgaagtgaagtctgttcggaagcgaggaggtttaatgccaccacaaactgt 868
Db 476 CGGCAGCACTGAAGTCAGTTTGCTGCTGGCACTGAAAGAGTCAATGGCACAGAGCTGT 535
Qy 869 ttgaagttgacacagacagctgtgagcgctgagctgcaaaagcgagttcttaaga 928
Db 536 TTGGNAGTGATACCGACAGCTGTGAGCGGTGGATGAACCTGCAAGAGTGAGTTTCTCAAGA 595
Qy 929 agtacaatgacaaggtgatgaatgacctgcccagctgcccctgtctctacccccactgagg 988
Db 596 A-TACATGAGTAAAGTAGCCACTGATCTACCCAGCTGCCCTGCTTTTACCCACCGAGG 654
Qy 989 tggcctacagca 1000
Db 655 TGGNCTACAGCA 666
```

Search completed: April 25, 2002, 11:39:10  
Job time: 5564 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 11:11:52 ; Search time 204.53 Seconds  
(without alignments)  
11710.241 Million cell updates/sec

Title: US-09-863-824-1

Perfect score: 1395

Sequence: 1 atggcgccctggcgccga.....aagaggccagggaattattaa 1395

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1393.8	99.9	1395	24	Human thrombospondin
2	875.8	62.8	936	24	Human thrombospondin
3	738.6	52.9	739	22	Human protein enco
4	554.6	39.8	790	21	Human OREF1686
5	495	35.5	495	24	Human thrombospondin
6	447.6	32.1	448	23	DNA encoding novel
7	325.2	23.3	2981	22	Human cDNA encoding
8	325.2	23.3	2981	22	Human proliferation
9	228.4	16.4	790	21	Human OREF1686

AA223967  
ID AAD23967 standard; cDNA; 1395 BP.  
AC AAD23967;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
DE Human thrombospondin-like protein cDNA #1.  
KW Human; thrombospondin-like protein; therapeutic; pharmacogenomic;  
KW drug screening; cosmetic; nutraceutical; anti-angiogenic; wound healing;  
KW endocrine; gene therapy; hyperthyroidism; hypothyroidism; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key  
CDS Location/Qualifiers  
FT 1..1395  
FT /\*tag= a  
FT /product= "Thrombospondin-like protein"  
FT /transl\_except= (pos:364..366, aa:Xaa)  
FT /note= "Xaa is Asn or Val"  
FT /transl\_except= (pos:535..537, aa:Xaa)  
FT /note= "Xaa is Lys or Glu"  
FT 364  
FT misc\_feature  
FT /\*tag= b  
FT /note= "This degenerate base represents A-G  
FT transition polymorphism"  
FT 365  
FT misc\_feature  
FT /\*tag= c  
FT /note= "This degenerate base represents A-T  
FT transversion polymorphism"

## ALIGNMENTS

RESULT 1

AA223967

ID AAD23967 standard; cDNA; 1395 BP.

AC AAD23967;

XX

XX

DT 26-MAR-2002 (first entry)

DE Human thrombospondin-like protein cDNA #1.

XX

XX

KW Human; thrombospondin-like protein; therapeutic; pharmacogenomic;

KW drug screening; cosmetic; nutraceutical; anti-angiogenic; wound healing;

KW endocrine; gene therapy; hyperthyroidism; hypothyroidism; ss.

XX

OS Homo sapiens.

XX

FH Key

CDS Location/Qualifiers

FT 1..1395

FT /\*tag= a

FT /product= "Thrombospondin-like protein"

FT /transl\_except= (pos:364..366, aa:Xaa)

FT /note= "Xaa is Asn or Val"

FT /transl\_except= (pos:535..537, aa:Xaa)

FT /note= "Xaa is Lys or Glu"

FT 364

FT misc\_feature

FT /\*tag= b

FT /note= "This degenerate base represents A-G

FT transition polymorphism"

FT 365

FT misc\_feature

FT /\*tag= c

FT /note= "This degenerate base represents A-T

FT transversion polymorphism"

DNA encoding novel  
DNA encoding novel  
Human cDNA encoding  
Human silent SNP c  
Human silent SNP c  
Sequence encoding  
Human silent SNP c  
Human silent SNP c  
Micromonospora DNA  
C899p determined c  
Consensus sequence  
Synthetic Bacillus  
B t maize optimise  
Synthetic Bacillus  
Maize optimised ge  
Maize-optimised ge  
B t maize optimise  
Human silent SNP c  
Human silent SNP c  
Zea mays DNA fragm  
Streptomyces prist  
HIV codon altered  
Micromonospora DNA  
Murine mahogany pr  
Murine mahogany pr  
Human mahogany pro  
Human mahogany pro  
Human mahogany pro  
Wild type (C57BL/6  
Gene encoding a su  
Rat TGF-beta bindi  
Truncated HSVgB ge  
Human heat shock p  
Herpes Simplex Vir  
HSV surface antige











XX New isolated nucleic acid molecule that encodes human  
PT thrombospondin-like protein, useful in therapeutic, diagnostic and  
PT pharmacogenomic applications, in drug screening and clinical trial  
PT monitoring  
XX  
XX Claim 1; Page 37; 39pp; English.  
XX  
XX The invention relates to nucleic acid molecule encoding human  
CC thrombospondin-like protein. The polynucleotide and polypeptide of  
CC the invention are useful in therapeutic, diagnostic and pharmacogenomic  
CC applications, in drug screening, clinical trial monitoring, and in  
CC cosmetic or nutraceutical applications. The polynucleotide is useful for  
CC identifying coding sequences and mapping a unique gene to a particular  
CC chromosome, to screen libraries, isolate clones, prepare cloning and  
CC sequencing templates, as hybridisation probes for screening libraries  
CC and assessing gene expression patterns, in microarrays and other assay  
CC formats to screen collections of genetic material from patients who  
CC have a particular medical condition, to identify mutations associated  
CC with a particular disease, in a diagnostic or prognostic assay, as  
CC antisense molecules, and as part of ribozymes and/or triple helix  
CC sequences that are useful for gene regulation, for detecting mutant  
CC proteins or inappropriately expressed proteins for the diagnosis of  
CC disease and in the molecular mutagenesis/evolution of proteins that  
CC are at least partially encoded by the polynucleotide. The  
CC thrombospondin-like protein or its modified processed form can  
CC be used as therapeutics e.g. anti-angiogenic agents, to promote wound  
CC healing and regulate endocrine functions. The polynucleotide  
CC can also be used in gene therapy and for treating disorders like  
CC hyperthyroidism and hypothyroidism. The present sequence is a cDNA  
XX encoding human thrombospondin-like protein.  
XX  
XX Sequence 495 BP; 126 A; 156 C; 139 G; 74 T; 0 other;

Query Match 35.5%; Score 495; DB 24; Length 495;  
Best Local Similarity 100.0%; Pred. No. 3.6e-102;  
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 901 atgagctcaaaagcgaagcttcttaaaagatcatatgcacaaaggtgatgaatgacctgcc 960  
Db 1 atgagctcaaaagcgaagcttcttaaaagatcatatgcacaaaggtgatgaatgacctgcc 60  
Qy 961 agctgccctgctctacccactgaggtggcctacagcagcgccgacatcttgacccg 1020  
Db 61 agctgccctgctctacccactgaggtggcctacagcagcgccgacatcttgacccg 120  
Qy 1021 atcaagcgaaggaacttcgctggaaggaacgacccagcgccgaaggaagctgagatc 1080  
Db 121 atcaagcgaaggaacttcgctggaaggaacgacccagcgccgaaggaagctgagatc 180  
Qy 1081 tacaagcccaactgccggtactgcctccgctccatgctgctccctggagagcaccacgctg 1140  
Db 181 tacaagcccaactgccggtactgcctccgctccatgctgctccctggagagcaccacgctg 240  
Qy 1141 gcggcacagcagctgtctacgacgacacacacacacacacacacacacacacacacac 1200  
Db 241 gcggcacagcagctgtctacgacgacacacacacacacacacacacacacacacacac 300  
Qy 1201 ggcagcccaactcatcagcagcagcttctccgagcagctccactacaaggtgagctc 1260  
Db 301 ggcagcccaactcatcagcagcagcttctccgagcagctccactacaaggtgagctc 360  
Qy 1261 ctgacctgattatctcaaggtgactggagcaggtataacgagggcccgccctcccaac 1320  
Db 361 ctgacctgattatctcaaggtgactggagcaggtataacgagggcccgccctcccaac 420  
Qy 1321 aacgacagaagtgcacagagagccctcgagcagaggtacatcaacagcttccaag 1380  
Db 421 aacgacagaagtgcacagagagccctcgagcagaggtacatcaacagcttccaag 480  
Qy 1381 gccagggaattata 1395  
|||||

Db 481 gccagggaattata 495  
RESULT 6  
AAS70533  
ID AAS70533 standard; cDNA; 448 BP.  
XX  
XX AAS70533;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #6337.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG06346.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 1; SEQ ID NO 6337; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 448 BP; 123 A; 122 C; 123 G; 80 T; 0 other;

Query Match 32.1%; Score 447.6; DB 23; Length 448;  
Best Local Similarity 99.8%; Pred. No. 1.7e-91;  
Matches 447; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 522 ggaccaggaactacagacagctacacagcagcagcagcagcagcagcagcagcagcagcag 581  
|||||  
Db 1 ggaccaggaactacagacagctacacagcagcagcagcagcagcagcagcagcagcagcag 60





05-APR-1999; 99US-0127728.  
30-MAR-2000; 2000US-0540763.  
(CURA-) CURAGEN CORP.  
Shimkets RA, Leach M;  
WPI; 2000-602362/57.  
P-PSDB; AAB41922.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 5; Page 2553; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
XX Sequence 790 BP; 181 A; 238 C; 212 G; 157 T; 2 other;  
SQ

[illegible]

[illegible]

[illegible]

RESULT	13	
AAI/7822		
ID	AAI/7822 standard; DNA; 51 BP.	
XX		
XX	AAI/7822;	
XX		
DT	09-NOV-2001 (first entry)	
XX		
DE	Human silent SNP containing nucleic acid SEQ.5663.	
XX		
XX	Human; single nucleotide polymorphism; SNP; genome; gene therapy;	
KW	protein therapy; vaccine; probe; diagnostic assay; detection;	
KW	quantitation; restorative therapy; polymorphic; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200140521-A2.	
XX		
PD	07-JUN-2001.	
XX		
PF	30-NOV-2000; 2000WO-US32758.	
XX		
PR	30-NOV-1999; 99US-0168138.	
PR	29-NOV-2000; 2000US-0726173.	
XX		
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Shimkets RA, Leach M;	
XX		
DR	WPI; 2001-356160/37.	
XX		
XX	Polymorphic nucleic acid sequences, useful in genetic testing and	
PT	therapy -	
PT		
XX	Claim 1; Page 2243; 2653pp; English.	
PS		

AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide sequences (I), which contain single nucleotide polymorphisms (SNPs). AAI53114 to AAI53329 represent peptides related to human polymorphic polynucleotide sequences. The sequences can be used in gene and protein therapy, and in vaccine production. (I) and the polypeptides encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate expression of polymorphic polypeptides. For example, (I) may be used to treat disorders by rectifying mutations or deletions in a patient's genome that affect the activity of polypeptides by expressing inactive proteins or to supplement the patients own production of polypeptide. Additionally, (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids

CC in samples, and therefore which patients may be in need of restorative  
CC therapy. The polypeptides encoded by (I) may be used as antigens in the  
CC production of antibodies specific for polymorphic polypeptides. The  
CC antibodies may also be used to down regulate expression and activity.  
CC The antibodies may also be used as diagnostic agents for detecting the  
CC presence of polymorphic polypeptides in samples.  
XX  
XX  
S0 Sequence 51 BP: 17 A; 19 C; 11 G; 4 T; 0 other;

Query Match	3.7%	Score 51	DB 22	Length 51
Best Local Similarity	100.0%	Pred. No. 0.021		
Matches	51	Conservative	0	Mismatches 0
		Indels	0	Gaps 0

Qy 191 ccaaagaagcacaaggagcatctgaccaccaggctgcacaccaacccct 241  
|||||  
pb 1 ccaaagaagcacaaggagcatctgaccaccaggctgcacaccaacccct 51

RESULT	14	
AAI78724		
ID	AAI78724	standard; DNA; 51 BP.
XX	XX	
XX	AAI78724;	
XX	XX	
XX	XX	
DT	DT	
XX	09-NOV-2001	(first entry)
DE	Human	silent SNP containing nucleic acid SEQ.5665.
XX	XX	
XX	Human;	single nucleotide polymorphism; SNP; genome; gene therapy;
KW	KW	protein therapy; vaccine; probe; diagnostic assay; detection;
KW	KW	quantitation; restorative therapy; polymorphic; ds.
XX	XX	
OS	Homo sapiens.	
XX	XX	
PN	WO200140521-A2.	
XX	XX	
PD	07-JUN-2001.	
XX	XX	
PF	30-NOV-2000;	2000WO-US32758.
XX	XX	
PR	30-NOV-1999;	99US-0168138.
PR	29-NOV-2000;	2000US-0728173.
XX	XX	
PA	(CURA-)	CURAGEN CORP.
XX	XX	
PI	Shimkets RA,	Leach M;
XX	XX	
XX	WPI;	2001-356160/37.
DR	XX	
XX	Polymorphic	nucleic acid sequences, useful in genetic testing and
PT	therapy -	
PT	therapy	
PS	Claim 1;	Page 2244; 2653pp; English.

AA173060 to AA179867 represent isolated human polymorphic polynucleotide sequences (I), which contain single nucleotide polymorphisms (SNPs). AA553114 to AA553329 represent peptides related to human polymorphic polynucleotide sequences. The sequences can be used in gene and protein therapy, and in vaccine production. (I) and the polypeptides encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate expression of polymorphic polypeptides. For example, (I) may be used to treat disorders by rectifying mutations or deletions in a patient's genome that affect the activity of polypeptides by expressing inactive proteins or to supplement the patients own production of polypeptide. Additionally, (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The polypeptides encoded by (I) may be used as antigens in the production of antibodies specific for polymorphic polypeptides. The antibodies may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the



CC presence of polymorphic polypeptides in samples.  
XX  
SQ Sequence 51 BP; 17 A; 18 C; 11 G; 5 T; 0 other;

```

Query Match      3.7%; Score 51; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 51; Conservative 0; Mismatches 0; Indels
Oy 192 caaagaagcaccaggagcatctggaccaccaggctgcacacacccct 242
Db 1 caaagaagcaccaggagcatctggaccaccaggctgcacacacccct 51

```

**RESULT** 15

AAQ63293  
ID AAQ63293 standard: DNA: 12588 BP.

AC AAQ63293;

DT 04-JAN-1995 (first entry)

DE Sequence encoding mannuronan C-5-epimerase genes.

Mannuronan; epimerase; alginates; immunostimulant;  
KW  
cell immobilisation; Azotobacter vinelandii; guluronic acid;  
KW  
mannuronic acid; ss.  
KW

OS Azotobacter vinelandii strain E.

.....	Key	Location/Qualifiers
FH		

```
FT      230.11551
/*tag= a
/product= Epimerase 4.
```

```
FT
000
00007110430
/*tag= b
```

FT	CDS
6702.9695	

FT /product= Epimerase 2.

```
FT
000
/*tag= d
```

```

FT      "The specification states the sequence
FT      length to be 12411 base pairs but the
FT      sequence reproduced in the specification
FT      is 12588 base pairs. The specification
FT      states that the first part of epimerase
FT      terminates at position 12411 although
FT      sequence there is no stop codon found
FT      position and the given nucleotide sequence
FT      is decoded in its entirety to give the
FT      corresponding amino acid sequence."
FT

```

PN WO9409124-A.

PD 28-APR-1994.

PF 08-OCT-1993; 93WO-NO00151.

PR 08-OCT-1992: 92GB-0021163.

PA (NOBI-) NOBIPOL NOBIPOLS FORSKNINGSSTIFTELSE.

FA (FRON-) FRONOVA BIOFOLIMER AS:  
XX

FI  
XX  
Lillesvåg n, Lillesvåg b, valla s, skjak-braek g,

DR WF1, 1934-131310/181  
DR P-PSDB: AAR54628, AAR

PT New nucleic acid encoding mannuronan C-5-epimerase

PT or M content, useful e.g. as immunostimulants or for cell



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 10:07:51 ; Search time 1729.22 Seconds  
(without alignments)  
16881.900 Million cell updates/sec

Title: US-09-863-824-1  
Perfect score: 1395  
Sequence: 1 atgggtgccttgcggccga.....aagagggccagggaataataa 1395

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: gb\_wa.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	1393.8	99.9	1395	6	AX351002	AX351002 Sequence
2	875.8	62.8	936	6	AX351006	AX351006 Sequence
3	519	37.2	84122	9	AL133463	Human DNA
4	495	35.5	495	6	AX351004	AX351004 Sequence
5	457	32.8	1617	9	BC017997	Homo sapi
6	326.8	23.4	3052	9	AK056709	Homo sapi
7	325.2	23.3	2981	6	AX136231	Sequence
8	268.8	19.3	231484	9	AF111168	Homo sapi
9	183.2	13.1	107057	9	HSJ107712	Human DNA
10	164.2	11.8	523	11	HS107712S	AL110084 H. sapiens
11	55.2	4.0	125020	9	AF429315	Homo sapi
12	54.2	3.9	45519	1	FVBPOAD2A	D26094 Flavobacter
13	52.8	3.8	1347	9	AF257167	AF257167 Homo sapi
14	52.4	3.8	7585	1	SCY14206	Y14206 Streptomyce
15	52.4	3.8	42816	1	SCC88	AL139298 Streptomy
16	52.2	3.7	34182	1	SC111	AL096849 Streptomy
17	51.8	3.7	42210	1	SC1C2	AL031124 Streptomy
18	51	3.7	51	6	AX162335	Sequence
19	51	3.7	51	6	AX162337	Sequence
20	50.6	3.6	3778	1	AVIALGEA	L39013 Azotobacter
21	50.6	3.6	15759	1	AVIALGEB	L39096 Azotobacter
22	50.2	3.6	51	6	AX162339	Sequence
23	50.2	3.6	51	6	AX162341	Sequence
24	50.2	3.6	1839	6	AX196078	Sequence
25	50.2	3.6	11905	1	APU33059	U33059 Actinosynne
26	50.2	3.6	109519	6	AX195929	Sequence
27	50.2	3.6	161970	8	AP003221	AP003221 Oryza sat
28	50	3.6	37245	1	SC5F2A	AL049587 Streptomy
29	49.8	3.6	6040	1	CTU32622	U32622 Comamonas t
30	49.6	3.6	3108	9	HUMIMUCB	M94132 Human mucin
31	49.6	3.6	15720	6	AX193491	Sequence
32	49.6	3.6	15720	6	AX330186	Sequence
33	49.6	3.6	15720	9	HUMMUC2X	L21998 Homo sapien
34	49.4	3.5	10959	1	AE004953	AE004953 Pseudomon
35	49.4	3.5	125020	9	AF429315	AF429315 Homo sapi
36	49	3.5	1149	6	AX089208	Sequence
37	48.8	3.5	1158	6	AR111753	Sequence
38	48.8	3.5	1158	6	AX089125	Sequence
39	48.8	3.5	152396	8	AP002971	AP002971 Oryza sat
40	48.6	3.5	51	6	AX162340	Sequence
41	48.6	3.5	51	6	AX162342	Sequence
42	48.4	3.5	592	8	WHTPH288B	D37943 Wheat mRNA
43	48.4	3.5	158186	2	AC096212	AC096212 Rattus no
44	48.2	3.5	137332	8	AP002820	AP002820 Oryza sat
45	48	3.4	2888	6	A48323	A48323 Sequence 1

## ALIGNMENTS

RESULT 1	AX351002	Sequence 1 from Patent WO0190179.	1395 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX351002	AX351002				
DEFINITION	AX351002	AX351002				
ACCESSION	AX351002	AX351002				
VERSION	AX351002.1	GI:18616372				
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (sites)					
AUTHORS	Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., AbuIn,A., zambrowicz,B. and Sands,A.T.					
TITLE	Novel human thrombospondin-like proteins and polynucleotides encoding the same					
JOURNAL	Patent: WO 0190179-A 1 29-NOV-2001;					
FEATURES	Lexicon Genetics Incorporated (US)					
source	Location/Qualifiers					
	1..1395					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
BASE COUNT	345 a	431 c	232 t			3 others



Qy 181 ttctctctccaagaagcaccagagagcatctggaccaccaggtgcacacaccc 240  
 Db 181 TTTTCTCTCTCCAAAGACACCAAGGAGGATCTGGACCAACAGGCTGCACACCAACC 240  
 Qy 241 ttccccagacgcgattccagacagagacggggcaccccttcaatgcaaaagagattcccc 300  
 Db 241 TTCCCCAGACCGGATTCGCAAGAGACGGGGCACCCCTTCATTGCAAGAGATTTCGCC 300  
 Qy 301 agatccttctccttgatctaccacactttccagatctttccaaagtgtatcaatggg 360  
 Db 301 AGATCCTTTCTCTTGATCTACCAAACTTTCCAGATCTTTCCAAAGCTGATCAATGGG 360  
 Qy 361 cagrtwccaaatccagggtcaccatagagtggtgcagcgtctctgactctgaagcagat 420  
 Db 361 CAGRWTCCAAATATCCAGGTCAACATAGAGTGTGTCAGCGTCTCTGACTCTGAAGCAGAT 420  
 Qy 421 aaagatcagatccggagataaagccagctgggtcagtcctccatcccccgactggcgcc 480  
 Db 421 AAAGATCAGCATCCGGAGAAATGAAGCCAGCTGGTCACTGCCATCCCGGACTGGCGGCC 480  
 Qy 481 tggtagcagaggtccctgtccttgccagggcaaacagcggggagccaggaactacragtac 540  
 Db 481 TGGTGGCAGAGTCCCTGTCTCTGGCCAGGGCAACACAGCGGGACCAAGGACTACRAGTAC 540  
 Qy 541 gacagtacctcagacgacagcaacttctcaacccccccaggggtgggacacacacagcc 600  
 Db 541 GACAGTACCTCAGACGACAGCAACTTCTCAACCCCCCAGGGGTGGGACCATACAGCC 600  
 Qy 601 ccaggccacgcggaattttgaaaccaaagatcagccagaatatgatccacagatggcgag 660  
 Db 601 CCAGGCCACCGGACTTTTGAACCAAAAGATCAGCCAGAATATGATTCACAGATGGCGAG 660  
 Qy 661 ggtcactggaactctgtctgtctgcagcgtccactcgtcggaagcggcaacacagaacgg 720  
 Db 661 GGTGACTGGAGTCTGTGTCTGTCTGCAGCGTCACTTCCGGGAACGGCAACCAAGAAACGG 720  
 Qy 721 acccggctcttggtgcagcgtgcactgcacagacagaatcagagacctgtgaccgtccaaac 780  
 Db 721 ACCCGGTCTTGTGCTACGCGTGCACATGCAACAGAAATCGAGGACCTGTGACCGTCCAAAC 780  
 Qy 781 tgcacagaattgagacacacttttagacagctcaccacgaagtgaagtctgcttgcggga 840  
 Db 781 TGCCACAGGAATTAAGACACACTTTAGGACAGCTGCCACCAAGTGAAGTCTGCTGCGGGA 840  
 Qy 841 agcagagagtttaagtcacacacaaactgtttgaagtgg 877  
 Db 841 AGCAGAGGTTAATGCCACCAACTGTTGAAGTTG 877

## RESULT 3

AL133463 Human DNA sequence from clone Rp11-149118 on chromosome 20.  
 LOCUS Contains the 3' end of a novel gene, ESTs, STSs and GSSs, complete sequence.  
 DEFINITION AL133463  
 ACCESSION AL133463  
 VERSION AL133463.16 GI:10443352  
 KEYWORDS HTS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 84122)  
 AUTHORS Wilson, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 COMMENT requests: clonerequest@sanger.ac.uk  
 On Oct 1, 2000 this sequence version replaced gi:10178396.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr20  
 IMPORTANT: This sequence is not the entire insert of clone Rp11-149118 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP4-585114 is at 84023 in this sequence. The true right end of clone RP5-107712 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. Rp11-149118 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6.

## FEATURES

source 1..84122  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /clone="Rp11-149118"  
 /clone\_lib="RPCI-11.1"  
 join(<1..159,9050..9314,17956..18099,21790..21879,28358..30067)  
 /gene="bA149118.1"  
 /note="match: ESTs: Em:A1191256 Em:R44763 Em:W86073 Em:A1082824 Em:AT886394 Em:R52250 Em:AV708785 Em:W07772 Em:AV660737 Em:BF302850 Em:AW426875 Em:W86257 Em:BF302850 Em:AW873508 Em:BE287313 Em:AW495535 Em:AW435648 Em:AW435639"  
 /product="bA149118.1 (novel protein)"  
 /evidence="not\_experimental"  
 1..30067  
 /gene="bA149118.1"  
 join(<1..159,9050..9314,17956..18099,21790..21879,28358..30067)  
 /gene="bA149118.1"  
 /note="continued from dJ107712.1 in Em:AL050320"  
 match: proteins: Tr:O95432 Tr:O9TT55"  
 /codon\_start=1  
 /evidence="not\_experimental"  
 /product="bA149118.1 (novel protein)"  
 /protein\_id="CAC16127.2"  
 /db\_xref="GI:13374941"  
 /translation="HOAAHOPFPRFRQETGHPSLQDPFRSFLDLNPNPDLKAD INGNPNIOVTIEVVDGPDSEADKQHPKPSVSPDMRAWQWQSLARNSGD QDIKIDSDSNFLNPPRGWDHPTAGHRTETKQPEYDSTGEGDWLSVCSVTC GNGNKRTRSCGYACTATCDPRNCPGIEDTFRFAATEVSLLAGSEEFNATKLF EVDTSCEWMSCKSEFLKYMVMNDLPCSPCYPTPEYATADIFDKRKDFRMK DASGPKLEIYKPTARYCIRSMLSLESTTLAAHCCYGDNMQLITGKGAGTPNLIS TEFSAEHLHYKVDVLPWTICKDWSRYNEARPNNGOKCTSPSDEDYIKQFQEAERY complement(1..81)  
 /note="match: STS: Em:HS1077125"  
 437..504  
 /note="34 copies 2 mer aa 66% conserved"  
 445..510  
 /note="2 copies 33 mer 93% conserved"  
 885..1046  
 /note="MIR repeat: matches 35..189 of consensus"  
 1202..1494  
 /note="AluX repeat: matches 1..291 of consensus"

## misc\_feature

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

repeat_region	2642.. 2951	/note="AluJo repeat: matches 1. .311 of consensus"
repeat_region	3460.. 3507	/note="24 copies 2 mer tg 79% conserved"
repeat_region	4065.. 4106	/note="21 copies 2 mer cc 76% conserved"
misc_feature	complement(4442..5302)	
repeat_region	4908.. 5154	/note="match: GSS: Em:AQ741486"
misc_feature	5180.. 5672	/note="MIR repeat: matches 34.. .238 of consensus"
repeat_region	7941.. 8067	/gene="ba149i18.1"
repeat_region	8078.. 8510	/note="MIR repeat: matches 1. .140 of consensus"
repeat_region	8538.. 8624	/note="MLT1H repeat: matches 81. .519 of consensus"
misc_feature	complement(8621..9044)	
misc_feature	complement(8758..9044)	
repeat_region	9887.. 10251	/note="match: GSS: Em:AQ133736"
repeat_region	10536.. 10673	/note="MLT1I repeat: matches 19. .371 of consensus"
repeat_region	10826.. 11052	/note="L2 repeat: matches 2563.. .2708 of consensus"
repeat_region	11277.. 11343	/note="MIR repeat: matches 2. .261 of consensus"
repeat_region	11344.. 11517	/note="MERSA repeat: matches 99. .166 of consensus"
repeat_region	11520.. 11698	/note="MERSA repeat: matches 1. .182 of consensus"
repeat_region	11773.. 11825	/note="MERSB repeat: matches 2. .178 of consensus"
repeat_region	12156.. 12212	/note="MERSA repeat: matches 136. .187 of consensus"
repeat_region	12248.. 12291	/note="MERSA repeat: matches 132. .189 of consensus"
repeat_region	12443.. 12595	/note="MERSA repeat: matches 9. .52 of consensus"
misc_feature	complement(13043..13439)	
repeat_region	13397.. 13552	/note="match: GSS: Em:B33790"
repeat_region	14382.. 14420	/note="MIR repeat: matches 35. .193 of consensus"
repeat_region	14421.. 14739	/note="MLT1H repeat: matches 514. .547 of consensus"
repeat_region	14740.. 15085	/note="AluJo repeat: matches 1. .310 of consensus"
repeat_region	16534.. 16892	/note="MLT1H repeat: matches 117. .514 of consensus"
repeat_region	16900.. 17163	/note="FHE1B repeat: matches 1. .364 of consensus"
repeat_region	17207.. 17360	/note="L2 repeat: matches 2491. .2747 of consensus"
repeat_region	18198.. 18489	/note="MLT1H repeat: matches 415. .568 of consensus"
repeat_region	18940.. 19047	/note="AluSg repeat: matches 13. .301 of consensus"
repeat_region	19018.. 19116	/note="MERS8C repeat: matches 1. .132 of consensus"
repeat_region	19103.. 19197	/note="3 copies 33 mer 82% conserved"
repeat_region	19785.. 20098	/note="MERS8C repeat: matches 119. .89 of consensus"
repeat_region	20136.. 20382	/note="AluJ repeat: matches 2. .306 of consensus"
repeat_region	21260.. 21460	/note="MIR repeat: matches 8. .261 of consensus"
repeat_region		/note="MIR repeat: matches 2. .219 of consensus"
repeat_region	22331.. 22644	/note="AluY repeat: matches 1. .302 of consensus"
repeat_region	23110.. 23419	/note="AluSg repeat: matches 3. .312 of consensus"
repeat_region	23996.. 24102	/note="LTR33 repeat: matches 415. .521 of consensus"
repeat_region	24208.. 24360	/note="MIR repeat: matches 38. .192 of consensus"
repeat_region	24986.. 25364	/note="LTR16B repeat: matches 1. .449 of consensus"
repeat_region	25661.. 25780	/note="LTR16C repeat: matches 1. .134 of consensus"
repeat_region	25792.. 26153	/note="MERS2B repeat: matches 272. .631 of consensus"
repeat_region	26154.. 26431	/note="AluSg repeat: matches 1. .278 of consensus"
repeat_region	26432.. 26701	/note="MERS2B repeat: matches 2. .272 of consensus"
repeat_region	26702.. 26928	/note="LTR16C repeat: matches 150. .387 of consensus"
repeat_region	27502.. 27831	/note="AluJb repeat: matches 1. .312 of consensus"
repeat_region	27890.. 27960	/note="L1MC4 repeat: matches 7716. .7789 of consensus"
repeat_region	27988.. 28097	/note="L1MC4 repeat: matches 7857. .7977 of consensus"
misc_feature	complement(29592..30009)	
polyA_signal	29986.. 29991	/note="match: STS: Em:G21248"
polyA_site	30009	/gene="ba149i18.1"
Query Match 37.2%; Score 519; DB 9; Length 84122;		
Best Local Similarity 100.0%; Pred. No. 4.2e-81;		
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy 877	gacacagacagctgtgagcgtgagctgagctgcaaaagcaggtcttaagaagtacatg 936	
Db 28357	GACACACACAGCTGTGAGCGCTGGATGAGTGCARAAAGCGAGTTCCTTAAGAAGTACATG 28416	
Qy 937	cacaagtgatgaatgacctgccagctgccctgtctctaccactgaggtggcctac 996	
Db 28417	CACAAGTGATGAATGACCTGCCAGTGCCTGTCTTACCCACTGAGGTGGCTAC 28476	
Qy 997	agacgcccacatcttcgaccgcatcaacgcaaggaacttcgctggaagcgcagc 1056	
Db 28477	AGCACGCCGACATCTTCGACCGCATCAAGCGCAAGGACTTCCGCTGGAAGGACGCCAGC 28536	
Qy 1057	gggcccaagagaagctggagattcaagcccaactgccggtactgcatccctccatg 1116	
Db 28537	GGGCCCCAAGGAGAGCTGGAGATCTACAGCCCACTGCCCGGTACTGCATCCGCTCCATG 28596	
Qy 1117	ctgtccctggagagcacacagctggcgacagcactgctgtctacgcgcaacaatgcag 1176	
Db 28597	CTGTCCCTGGAGACACACAGCTGGCGGCACAGCACTGCTGCTACGGCGCAACAATCGAG 28656	
Qy 1177	ctcatcaccaggggcaagggggcgggcgacgcccaacctcatcagcacgcaggttctccg 1236	
Db 28657	CTCATCACAGGGGCAAGGGGGCGGCACGCCCACTCATCAACACCCAGGAGTCTCCCGC 28716	
Qy 1237	gagctccactacaaggtggacgtctcctgcctggattatctgcaaggtgactgagcag 1296	
Db 28717	GAGTCCCACTACAGGTGGAGCTCTGCCCTGGATTATCTGCAAGGTGACTGGAGCAGG 28776	
Qy 1297	tatacagggccggcctcccaacaacggaagtcacagagagccctcgagcag 1356	
Db 28777	TATAACGAGGCCGCGCTCCCAACAACGACAGAGTGCACAGAGGCCCTCGACGAG 28836	
Qy 1357	gactacatcaagcaggttcccaagagccaggaatatataa 1395	
Db 28837	GACTACATCAAGCAGTTCCTCAAGAGGCCAGGAATATTAA 28875	

```
RESULT 4
AX351004
LOCUS AX351004 495 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO0190179.
ACCESSION AX351004
VERSION AX351004.1 GI:18616373
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., Abuin,A.,
Zambrowicz,B. and Sands,A.T.
TITLE Novel human thrombospondin-like proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0190179-A 3 29-NOV-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source
1..495
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 126 a 136 c 139 g 74 t
ORIGIN

Query Match 35.5%; Score 495; DB 6; Length 495;
Best Local Similarity 100.0%; Pred. No. 8.8e-77;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 atgagctgcacaaacgaggtctttaaagaagtacatgcacaaagtgatgaatgacctgcc 960
DB 1 ATGAGCTGCACAAACGAGGTCTTAAAGAGTACATGTCACAAAGGTGATGATGACCTGCC 60

QY 961 agctgcccctgctcctaccacctgagtggtgcttacagcagcggcgacatcttcgacgc 1020
DB 61 AGCTGCCCTGCTCCTACCCCACTGAGGTGGCTTACAGCAGCGCGGCGACATCTTCGACCG 120

QY 1021 atcaagcgaaggaattccctggaagagacgacgagcggcccaagaggaagactgagatc 1080
DB 121 ATCAAGCGCAAGGACTTCCCGTGGAAAGACGCCAGCGGGGCCCAAGGAAAGCTGGAGATC 180

QY 1081 tacaagccactgcccgtactgctacgcctccatgctgctccctggagagcaccacgctg 1140
DB 181 TACAAGCCCACTGCCGCTACTGCTCGCTCCATGCTGCTCCCTGGAGAGCACCACGCTG 240

QY 1141 gcggcacagcactgctgctacggcgcacaacatgcagctcatcacccagggcgcaaggcg 1200
DB 241 CGGCACAGCACTGCTGCTACGGCGGACAAACATGCAGCTCATCACAGGGGCGGCGGCG 300

QY 1201 ggcagcccaacctcatcacacagcttctccgagagctccactacaaggtgacgtc 1260
DB 301 GGCAGCCCAACCTCATCAGCACCGAGTCTTCCCGGAGCTCCACTACAAGGTGGAGCTC 360

QY 1261 ctgcccctggattatctcaagggtgactggagcaggtataacagagccggcctcccaac 1320
DB 361 CTGCCCTGGATTATCTGCAAGGTTGACTGGAGCAGGTATACGAGGCCGCGCTCCCAAC 420

QY 1321 aacggacagaagtgcacagagagccctcggacgaggaactacatcaagcagttccaagag 1380
DB 421 AACGGACAGAAGTGCACAGAGAGCCCCCTCGGACGAGGACTACATCAAGCAGATTCCAAGAG 480

QY 1381 gcaagggaattataa 1395
DB 481 GCCAGGGAATTATAA 495

RESULT 5
BC017997
LOCUS BC017997 1617 bp mRNA linear PRI 06-DEC-2001
DEFINITION Homo sapiens, clone IMAGE:4252124, mRNA, partial cds.
ACCESSION BC017997
```

```
BC017997.1 GI:17389973
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1617)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILLNL at: http://image.llnl.gov
Series: IRAL Plate: 35 Row: f Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
Location/Qualifiers
1..1617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4252124"
/tissue_type="Skin, melanotic melanoma, high MDR."
/clone_lib="NIH_MGC_62"
/lab_host="DH10B"
/notes="vector: pDNR-LIB"
<1..457
/codon_start=2
/product="Unknown (protein for IMAGE:4252124)"
/protein_id="AAH17997.1"
/db_xref="GI:17389974"
/translation="KVNQLPSCPSTPTEVAVSTADIFRIKRDPRWKDASGPKK
LEIYKPTARYCIKRSMSLSTTTAAQHCCYGDNNQLITRGKAGTPLNLISTEFSAEIH
YKVDVLPWIIICKDWSRYNEARPPNNQKCTESPSDEDIYKQFEAREY"

BASE COUNT 506 a 348 c 391 g 372 t
ORIGIN

Query Match 32.8%; Score 457; DB 9; Length 1617;
Best Local Similarity 100.0%; Pred. No. 3.5e-70;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 939 caagtgatgaatgaacctgccagctgccctgctcctaccacctgagtgacctacag 998
DB 1 CAAGGTGATGAATGACCTGCCAGCTGCCCTGCTCTCTACCCCACTGAGGTGGCTACAG 60

QY 999 cagggcgacatcttcgacccgcatcaagcgcaaggaactccgctggaggaagcgccagcg 1058
DB 61 CACGGCGCACATCTTCGACCCGCATCAAGCGCAAGGACTTCCGCTGGAGGACGCCGCGG 120

QY 1059 gcccaaggagaagctggagatctacaagccacctgcccggctactgcatccctccatgct 1118
DB 121 GCCCAAGGAGAAGCTGGAGATCTACAAGCCCACTGCCCGGTACTGCATCCGCTCCATGCT 180

QY 1119 gtccctggagagaccacgctggcgccacagcactgctgctacgagcgacacatgacgt 1178
DB 181 GTCCCTGGAGAGCACCAGCTGGCGGCACAGCACTGCTGCTACGGCGCAACAACATGACGT 240
```

```

QY 1179 catcaccagggaagggcgccgaccccaactcatcagcaccaggtctcccgagg 1238
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 241 CATCACCAGGGGCAAGGGGGGGGCGACGCCCAACCTATCAGCACCGGTTCTCGGGGA 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1239 gctccactacaaggtggaagctccctgctgattatctgcaaggtgactggagcagta 1298
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 301 GCTCCACATACAGGTGGAGGCTCTGCTGCTGATTATCTGCAAGGGTGACTGGAGCAGGTA 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1299 taacgagggccggcctcccaacaacgagcagaagtgcacagagccctcgagcagga 1358
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 361 TAACGAGGGCCGGCTCCCAACAGGACAGAGTGACAGAGGCCCTCGGACGAGGA 420
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1359 ctatcatcaagcagttcccaaggagggccaggggaatattaa 1395
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 421 CTATCATCAAGCAGTTCCCAAGAGGGCCAGGGAATATTAA 457
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||

RESULT 6
AK056709 Homo sapiens cDNA FLJ32147 fis, clone PLACE5000116.
LOCUS AK056709 Homo sapiens cDNA FLJ32147 fis, clone PLACE5000116.
DEFINITION AK056709 Homo sapiens cDNA FLJ32147 fis, clone PLACE5000116.
ACCESSION AK056709
VERSION AK056709.1 GI:16552189.
KEYWORDS oligo cappings; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACES
clone:PLACE5000116.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (sites)
Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
Isogai,T.
NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3052)
Isogai,T., Otsuki,T. and Sugiyama,T.
AUTHORS Direct Submission
TITLE Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
JOURNAL Genomics Laboratory; 1332-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, tel:81-438-52-3951, fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.
FEATURES
Source Location/Qualifiers
1. 3052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE5000116"
/tissue_type="placenta"
/clone_lib="PLACES"
/note="cloning vector: pME18SFL3"
BASE COUNT 622 a 946 c 910 g 574 t
ORIGIN

Query Match 23.4%; Score 326.8; DB 9; Length 3052;
Best Local Similarity 66.3%; Pred. No. 1.7e-47;
Matches 488; Conservative 0; Mismatches 242; Indels 6; Gaps 1;

QY 659 aggggtgactggagctctggtgtgtgtgctgagcggtcacctcggggaacggaacgaagaac 718
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1114 AGAAGAGTGGAGTCCCTGCTCTCCCTGCGTGGGAAGTGCAGCACTGGCAGCAGCAGCA 1173
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||

```

```

QY 719 ggacccgggttttggctacgctgctcactgcaacagaatacgaggacctgtgaccgtccaa 778
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1174 GGACTGGGCTGTGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1233
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 779 actgcccagaaattgaacacacttttaggacacactgccaccggaagtgaagtctgtcgcg 838
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1234 CTTGCTCTGGCACTGAAGACAAGACACCTTTGGGCTCCCTCCCTCCCTCCCTCCCTCCCT 1287
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 839 gaagcgaggagtttaattgccaccaaaactgtttgaagtgtgacacagacagctgtgagcgt 898
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1288 AGTCTCTGGCCCCCAATGCTACGACATGATCAAGATGTGGACAGCTGTGAGAAGT 1347
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 899 ggaatgagctcaaaagcaggttctttaaagaagtacatgcaaaaggtgataatgacctgc 958
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1348 GGCTGAACATGCAAGAGCGGCTTCTTAATCAAGTATCTGAGCCAGATGCTCGGGAGCTGC 1407
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 959 ccagctgcccctgctctaccaccactgaggtggcctacagcagcgccgacatcttcgacc 1018
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1408 CCAAGTCCCGGTGTGCTTACCCACTGTGAGGCCATGACAGCCCTGTGAGCTACAGGAGC 1467
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1019 gcatcaagcgcaaggacttccgctggaagacgcccagcgggcccaaggaggaagctggaga 1078
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1468 AGCACCAGGCGCGAGCTTCCGGTGGAGGATGCCAGTGGCCCTCGCGAGCGCTGGACA 1527
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1079 tctaaagcccactgcccggtaactgcatcgctccatgctgtccctggagagcaccagc 1138
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1528 TCTACCAGCCACGGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1587
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1139 tggcggcacagcactgctgtacggcgacacacatgcagctcatcaccaggggcaaggggg 1198
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1588 TGGCGCGCCAGCACTGCTGTATGACGAGGACAGCCGGCTGCTGACCCGTGGCAAGGGCG 1647
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1199 cgggacgcgcacaaactcatcagcaccaggttctccgcgagctccactacaaggtggagc 1258
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1648 CCGCATATGCCAACCTCATCAGCACCAGCTTCTCACCTTAAGCTGCACTTCAAGTTCGACA 1707
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1259 tctgcccctgattatctcaagggtgactgagcaggtataacgagggccggcctccca 1318
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1708 CGAGCCCTGTGATCTGTGAAGGGGACTGAGGCCCTCCAGCTGTCTCCCTCCCA 1767
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1319 acaacgacagaagtgcacagagagccctcgagcagggaggtacatcaagcagttccaag 1378
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 1768 ACAACGGCCGCGCTGTCACCCAGACAAACCCCTGGAGGAGGAGTACCTAGCACAGTTGAGG 1827
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1379 agggcagggaatatta 1394
||||| ||| || ||
Db 1828 AGGCCAAGGAGTACTA 1843
||||| ||| || ||

RESULT 7
AX136231 AX136231 2981 bp DNA linear PAT 30-MAY-2001
LOCUS Sequence 153 from Patent EP1067182.
DEFINITION AX136231
ACCESSION AX136231
VERSION AX136231.1 GI:14272639
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 2981)
Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
Hayashi,K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 153 10-JAN-2001;
Helix Research Institute (JP)
FEATURES Location/Qualifiers
source 1. 2981
/organism="Homo sapiens"
/db_xref="taxon:9606"
58. .1773
/note="unnamed protein product"
CDS

```



/codon_start=1									
/protein_id="CAC39758.1"									
/db_xref="GI:14272640"									
/translation="MRALDRAGLLVLLLAALLEALGLPVKKPRLRGPRGSLTR									
LAEVSADPPRLPKREEEAPLPRTHLOAEPHOHCWTVEPAAMTPGNTTPRTPEV									
TPLRLLEQLKPLGLASTLSTPNPDQASAPDRPLREEEARLLPRTHLOAEHOGH									
CWTVEAPALPPNPPTTQEVETPPLLEKLQPLVHATLSTPNPDNQVTKIVVEDP									
QAEVSIDLAEPSNPPTTQVLSWLPALWPLMGWDYGEKKDRAPGKEBEDEDEYP									
SEDIGEDDEKDEEEDQALWFGTNDMDQGLAPGWDFKDSYSDYEPQKWSF									
WSPGNCSTGKQORTPCQGTATETCTDLPSGPTKEDKDTLGLPSEEWKLLARN									
ATMDHDDVDSCFKLNSKDFLLKVLMSLRDLPCPCAYPLEAMDSPVSLQDEHG									
RSPFRWDAGSPRRLDIQYPTARFCLRSLMSGESSLSLAQHCCYDDEDSRLLTRGKGAG									
MPNLSTDSFSLKHFADFTPTPMLTGLGDSRLHVLPPNNRGACTDNPULEEYLAQLQ									
EAEKY"									
BASE COUNT	609	a	925	c	887	g	560	t	
ORIGIN									
Query Match	23.3%	Score	325.2	;	DB	6	Length	2981	
Best Local Similarity	66.2%	Pred. No.	3.3e-47						
Matches	487	Conservative	0	Mismatches	243	Indels	6	Gaps	1
Qy	659	agggtgactgagctctgctgtctgctgagcgtcacctcggaagcaggaacccaagaac	718						
Db	1043	AGAAGGAGTGGAGTCCCTGCTCCTCCGTCGAGCTGGGAACTGCAGCACTGGCAAGCAGAGA	1102						
Qy	719	ggacccggcttggctgactgcagctgcaacagagaatcgagagacctgtgacctgcaaa	778						
Db	1103	GGACTCGGCCCTGTGGCTATGTGGCTGCATGCCACCGAGACCCGTACCTGTGACCTGCCT	1162						
Qy	779	actgccaggaattgaagacaacttttaggacagctgcccacggaagtgagtctgttcg	838						
Db	1163	CCTCTCTGGCACTGAGGACAAGGACACCTTTGGGCCCTCCCC-----AGTGAGGAGTGA	1216						
Qy	839	gaecgaggagtttaatgcccacaaactgtttgaagttgacacagacagctgtgagcgt	898						
Db	1217	AGCTCTGGCCCGCAATGCTACGGACATGATGATCAAGATGTGGACAGCTGTGAGAAGT	1276						
Qy	899	ggatgagctgcaaaagcgagttcttaagaagtacatgcaacaggtgatgaatgacctgc	958						
Db	1277	GGCTGAAGTGAAGAGCGACTTCTTATCAAGTATCTGAGCCAGATGCTCGGGACCTGC	1336						
Qy	959	ccagctgcccctgctctaccccactgagtgggcctacagacagcgccgacatcttcgacc	1018						
Db	1337	CCAGCTGCCCTGTGCCCTACCCACTGGAGGCCATGGACAGGCCCTGTGAGCCTACAGGACG	1396						
Qy	1019	gcatcaagcgcaagaattccgctgggaagcagccagcgagggcccaagagagctggaga	1078						
Db	1397	AGCACAGGCGCCGAGCTCCGGTGGAGGATGCCATGGCCCTCCGCGAGCGCTGGACA	1456						
Qy	1079	tctaaagccactgcccggtaactgcatcgctccatgctgctccctggagagcaccacgc	1138						
Db	1457	TCTACAGCCACCGCGCGCTTCTGCCCTGCTTCATGCTGCTCTGGGGAGAGCAGCACAC	1516						
Qy	1139	tggcggcacagcactgtgctacggcgcaacaatgtagctctcatcaccagggggcaagggg	1198						
Db	1517	TGGCGGCCACAGCACTGTGCTATGACGAGGACAGCCGGCTGCTACCCCGTGGCAAGGCG	1576						
Qy	1199	cggcagcgcccaactcatcagacaccgagttctccgcgagagctcactacaagtgagac	1258						
Db	1577	CCGGCATGCCCCAACTCATCAGCACCACCTTCTCACCATAAGCTGCATTCAGTTTCACA	1636						
Qy	1259	tctgctcctgattatctgcaagggtgactggagcaggtataacagagcccgccctccca	1318						
Db	1637	CGAGCCCTGGATCTGTGCNAGAGGGGAGCTGGAGCCCGCTCCACGCTGTGCTCCTCCCA	1696						
Qy	1319	acaacgacagaagtcacacagagagccctctcgagcagagactacataaagcagttccaa	1378						
Db	1697	ACAACGGCCGAGCCTGCACCGACAAACCCCTGGAGGAGGAGTACCTAGCACAGTTGCAGG	1756						
Qy	1379	aggccagggaatatta	1394						
Db	1757	AGCCAAAGGAGTACTA	1772						

```

/rpt_family="5S"
complement(6715..6815)
/note="low quality data"
complement(6914..7458)
/rpt_family="LINE2"
7523..7707
/rpt_family="L1MB6"
7708..8007
/rpt_family="AluSx"
8233..8305
/rpt_family="(TA)n"
9186..9230
/rpt_family="(CA)n"
complement(join(9248..9470,10060..10240,16529..16722,
25330..25484))
/note="Intron-exon boundaries defined in relation to an
EST contigs that includes AA397580, A1138314, AA399545,
AA86184, all from testis. The closest match in BLASTX is
to E. coli pyruvate-5-carboxylate reductase."
/codon_start=1
/product="unknown"
/protein_id="AAD09625.1"
/db_xref="GI:4186186"
/translation="MDMLQDLESLOFEYGPEDRWLYQGRSLMIEACAHATFF
CKLYNLRASLNKNKNSRHSLSGNSATPEEFKVGILGGHGLQLAGTLQGLPIP
AESLRITRPEPTLQELKIGKFCYHNADLVSWADVFCLCLPSQLPNCVETIYSL
EKASIVSFVAATPLPRKLLNHTNLRPOYQYDSDSVSWGANKGVIAALDPTIL
QATCPYPAGHSGSYFWHSSEYFPVGPLLSY"
complement(9631..9708)
/rpt_family="LINE2"
complement(10461..10762)
/rpt_family="AluSx"
complement(10804..11077)
/rpt_family="AluJb"
complement(11156..11306)
/rpt_family="MIR"
complement(11493..11615)
/rpt_family="FLAM_C"
11800..12161
/rpt_family="RHEIC"
complement(12406..12695)
/rpt_family="AluSx"
12746..12920
/rpt_family="BC200"
12921..13217
/rpt_family="AluSx"
13236..13751
/rpt_family="L1ME1"
13783..14084
/rpt_family="AluSx"
14085..14403
/rpt_family="L1ME2"
14405..14717
/rpt_family="AluSx"
14795..14857
complement(14884..14905)
/rpt_family="AT-rich"
complement(15106..15393)
/rpt_family="AluSx"
complement(15606..15818)
/rpt_family="L1ME2"
complement(15835..16131)
/rpt_family="AluSc"
complement(16775..17073)
/rpt_family="AluY"
17491..17777
/rpt_family="AluJo"
17778..17804
/rpt_family="(CA)n"
complement(17815..17882)
/rpt_family="L1MB2"
17883..18175

```

```

repeat_region
18177..18251
/rpt_family="(TAGA)n"
complement(18244..18306)
/rpt_family="L1MB1"
repeat_region
complement(18310..18609)
/rpt_family="AluSx"
repeat_region
complement(18624..18771)
/rpt_family="L1MB3"
repeat_region
complement(18765..18847)
/rpt_family="L1"
18858..19501
/rpt_family="SVA"
complement(19503..20417)
/rpt_family="L1"
complement(20104..20625)
/rpt_family="L143_Send"
repeat_region
complement(20724..21029)
/rpt_family="AluJb"
repeat_region
complement(21030..21122)
/rpt_family="(TAAA)n"
complement(21125..21423)
/rpt_family="AluJo"
repeat_region
complement(21486..21662)
/rpt_family="L191_Send"
repeat_region
complement(21664..21966)
/rpt_family="AluSx"
repeat_region
complement(22042..22342)
/rpt_family="AluY"
complement(22350..22937)
/rpt_family="L143_Send"
repeat_region
complement(22975..23090)
/rpt_family="FLAM_C"
repeat_region
complement(23164..23287)
/rpt_family="L143_Send"
repeat_region
complement(26429..26704)
/rpt_family="L1MB5"
26766..27080
/rpt_family="AluSx"
complement(28232..28432)
/rpt_family="MER20"
28528..28689
/rpt_family="AluJb"
28692..28984
/rpt_family="AluSx"
complement(29016..29108)
/rpt_family="MIR"
complement(join(30211..30231,30965..31069,31601..31690,
32276..32362,36437..36526,36857..36898,37854..37988,
38439..38514,40403..40476,43661..43688,45155..45257,
45714..45747,46844..46936,51089..51145,51888..51952,
53782..53928,55894..55996,56605..56697))
/note="Intron-exon boundaries defined by an EST contig
that includes W33552, W80307, AA780270, AA78570, R53310,
and AA437679. There is a short form of this coding
sequence that terminates at position 188433, based on EST
AA780270. The closest similarity by BLASTX is to a C.
elegans hypothetical protein."

```

```

Query Match      19.3%; Score 268.8; DB 9; Length 231464;
Best Local Similarity 70.3%; Pred. No. 1.7e-37;
Matches 360; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy  893  gacagctgtgagcgtgagctgcaaaagcaggttcttaagaagatcatcaccaag 942
      |||||
Db  78702  GACAGCTGTGAGAGTGGCTGAACGTGAAGAGCGGCTTCCCTAATCAAGTATCTGAGCCAG 78643
      |||||

Qy  943  gtgatgaatgacctgccagctgccctgtctctaccaccactgaggtgacctacagcag 1002
      |||||
Db  78642  ATGCTGGGGAGCTGCCAGCTGCCGCTGCTTACCACCTGGAGGCCATGGACAGCCCT 78583
      |||||

Qy  1003  gccgacatcttcgaccgcacatcaagcgcaaggaactccgctggaagcagcagcggccc 1062
      |||||

```

Db 78582 GTGAGCTACAGGACGACACAGGCGCCGAGCTTCGGTGGAGGATGCCAGTGGCCCT 78523

QY 1063 aaggagaagctgagatctacaagccactgcccgggtactgcatcgcgctccatgctgctc 1122  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78522 CGGAGCGCTGGACATCTACAGCCACGCGCGCTTCTGCCTCGCTTCCATGCTGCT 78463

QY 1123 ctggagagaccacgtctgctggcgccacgactgtgtctacgctggcgacacatgcagctcatc 1182  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78462 GGGAGAGACGACACTGTGGCGCCGACGACTGTGCTATGACGAGGACACGCGCTGCTG 78403

QY 1183 accaggggcaaggggcgccgaccccaacctcatcagaccgagttctccgagagctc 1242  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78402 ACCCGTGGCAAGGCGCGCGCATGCCCCAACCTCATCAGCACCCGACTTCTCACCTAAGCTG 78343

QY 1243 cactacaagtgagctcctgcctggattatctgcaaggtgactggagcaggtataac 1302  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78342 CACTTCAAGTTCGACAGCAGCGCCTGTGATCTGCAAGGGGACTGGAGCCGCTCCAC 78283

QY 1303 gaggcccgccctcccaacacgacgaagtgccacagagagccctcgagcagagactac 1362  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78282 GCTGTGCTCCTCCCAACAACGCGCGAGCCTGCACCGACAAACCCCTGGAGGAGGTAC 78223

QY 1363 atcaagcagttccaagaggccagggaattatta 1394  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78222 CTAGCACAGTTGCAGGAGGCCAAGGAGTACTA 78191

RESULT 9

HSJ107712 107057 bp DNA linear PRI 04-APR-2001

LOCUS Human DNA sequence from clone RP5-107712 on chromosome 20. Contains

DEFINITION part of a novel gene, ESTs, STSs, GSSs and a CpG island, complete

sequence.

ACCESSION AL050320.19 GI:7263998

VERSION HTG: CpG island.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 107057)

TITLE Direct Submission

JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

COMMENT On Mar 19, 2000 this sequence version replaced gi:6523711.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 20, constructed by the Sanger Centre Chromosome 20

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr20

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. RP5-107712 is from

the library RPCI-5 constructed by the group of Pieter de Jong. For

further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone  
RP5-107712 it may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true right end of clone RP5-107712 is at 107057 in this  
sequence. The true right end of clone RP4-718P11 is at 100 in this  
sequence.

## FEATURES

source

Location/Qualifiers

1..107057  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone="RP5-107712"  
/clone\_lib="RPCI-5"  
286..490  
/note="LIP4 repeat: matches 5942..6146 of consensus"  
1467..1494  
/note="14 copies 2 mer ac 8% conserved"  
1869..2354  
/note="match: STS: Em:AA159919"  
2112..2748  
/note="match: STS: Em:AA005104"  
2399..2464  
/note="MER5A repeat: matches 41..105 of consensus"  
2513..2619  
/note="match: STS: Em:HSPF04E9 Em:294369"  
complement(2651..3124)  
/note="match: STS: Em:G29540 Em:T98667"  
complement(2856..3124)  
/note="match: STS: Em:G25883"  
3615..3653  
/note="3 copies 13 mer 94% conserved"  
4108..4247  
/note="MER5A repeat: matches 44..184 of consensus"  
complement(4488..5147)  
/note="match: GSS: Em:AQ262066"  
4541..4855  
/note="AluSc repeat: matches 1..309 of consensus"  
complement(4979..5508)  
/note="match: GSS: Em:AQ664680"  
complement(4996..5415)  
/note="match: STS: Em:N30188"  
5315..5364  
/note="25 copies 2 mer tg 84% conserved"  
5318..5365  
/note="12 copies 4 mer gtgt 85% conserved"  
5489..5566  
/note="LIR16A repeat: matches 378..450 of consensus"  
5567..5760  
/note="MER63A repeat: matches 1..209 of consensus"  
5761..6016  
/note="LIR16A repeat: matches 95..378 of consensus"  
6044..6288  
/note="LIMC5 repeat: matches 7695..7937 of consensus"  
7638..8163  
/note="match: GSS: Em:B52469"  
7665..8134  
/note="match: GSS: Em:AQ667183"  
8270..8512  
/note="HAL1 repeat: matches 162..421 of consensus"  
8329..8462  
/note="67 copies 2 mer aa 59% conserved"  
8970..9273  
/note="HAL1 repeat: matches 680..997 of consensus"  
10789..11086  
/note="AluJb repeat: matches 1..297 of consensus"  
11170..11523  
/note="THE1B repeat: matches 1..364 of consensus"  
11578..11752  
/note="MER86 repeat: matches 2..195 of consensus"  
complement(11947..12544)  
/note="match: STS: Em:AL110056 Em:HS75707T"  
12307..12526  
/note="LIMEC repeat: matches 1192..1424 of consensus"





Kato, K.  
 Direct Submission  
 Submitted (03-DEC-1993) Ko Kato, Osaka University, Department of  
 Biotechnology; 2-1 Yamada-oka, Suita, Osaka 565, Japan  
 (E-mail: tyomoe@dbj.nig.ac.jp, Tel: 06-877-5111 (ex. 4372),  
 Fax: 06-876-2250)  
 2 (sites)  
 Okada, H., Negoro, S., Kimura, H. and Nakamura, S.  
 Evolutionary adaptation of plasmid-encoded enzymes for degrading  
 nylon oligomers  
 JOURNAL  
 MEDLINE  
 Nature 306 (5939), 203-206 (1983)  
 3 (sites)  
 Tsuchiya, K., Fukuyama, S., Kanzaki, N., Kanagawa, K., Negoro, S. and  
 Okada, H.  
 High homology between 6-aminohexanoate-cyclic-dimer hydrolases of  
 Flavobacterium and Pseudomonas strains  
 J. Bacteriol. 171 (6), 3187-3191 (1989)  
 89255082  
 4 (sites)  
 Negoro, S., Kakudo, S., Urabe, I. and Okada, H.  
 A new nylon oligomer degradation gene (nylC) on plasmid pOAD2 from  
 a Flavobacterium sp  
 J. Bacteriol. 174 (24), 7948-7953 (1992)  
 93094123  
 5 (sites)  
 Kato, K., Ohtsuki, K., Koda, Y., Maekawa, T., Yomo, T., Negoro, S. and  
 Urabe, I.  
 A plasmid encoding enzymes for nylon oligomer degradation:  
 nucleotide sequence and analysis of pOAD2  
 Microbiology 141 (Pt 10), 2585-2590 (1995)  
 96036217  
 6 (bases 1 to 45519)  
 Kato, K., Ohtsuki, K., Koda, Y., Maekawa, T., Yomo, T., Negoro, S. and  
 Urabe, I.  
 Structural analysis of nylon oligomer degradative plasmid poad2:  
 whole nucleotide sequence of poad2  
 Unpublished (1993)  
 7 (sites)  
 Kato, K., Ohtsuki, K., Mitsuda, H., Yomo, T., Negoro, S. and Urabe, I.  
 Insertion sequence is6100 on nylon oligomer degradative plasmid  
 poad2  
 Unpublished (1994)  
 Location/Qualifiers  
 1..45519  
 /organism="Flavobacterium sp."  
 /plasmid="plasmid pOAD2"  
 /strain="KI723T1"  
 /db\_xref="taxon:239"  
 1..859  
 /standard\_name="IS6100"  
 /note="RS-IA repeat sequence; The right part of IS6100 in  
 the RS-IA region; insertion sequence"  
 /citation=[6]  
 complement(2811..3989)  
 /gene="nylB"  
 complement(2811..3989)  
 /gene="nylB"  
 /standard\_name="EII"  
 /EC\_number="3.5.1.46"  
 /note="RS-IA repeated sequence; Sequence homologous to  
 RS-IB region; The EII enzyme is 100 times more active  
 toward the substrate than the EII' enzyme."  
 /citation=[2]  
 /codon\_start=1  
 /transl\_table=11  
 /label=RS-IA  
 /product="6-aminohexanoate-dimer hydrolase"  
 /protein\_id="BAA05087.1"  
 /db\_xref="GI:488340"  
 /translation="MNAKSTGHPARYPGAAGEPTLDSWQEPHNWFAHLGEMV  
 TAAVSRPAPAEVRLDLEQRLTECTDAFLVLRGEVLAAYRAGF  
 APDRHLLMSVSKSGCTVWGLIDEGRIDPAQVTEVPELAGSVLDQVLDLDM

QISIDYNEDYVDFASEVQTHDRSAGWTRRGDPADTYEELTTLRGDGGTGEYCSEA  
 NTVLAWLIVERTVGLRYVEALSTYLMKADLDRDATITVDQTGFGFRANGVYSTARDL  
 ARVGRMLDGGVAGPGRVVSQGWVESYLAGGSRDMDTEGTFSAFPEGSTYRWMCCTG  
 NERGNVSGIGIHQNLWLDPRDTSVIVKLSSWPPDTRHWHGLQSGILLDVSRAILDV  
 complement(3997..4000)  
 /note="Shine-Dalgarno sequence"  
 complement(4602..5669)  
 /gene="nylC"  
 complement(4602..5669)  
 /gene="nylC"  
 /standard\_name="EIII"  
 /citation=[4]  
 /codon\_start=1  
 /transl\_table=11  
 /label=EIII  
 /product="endo-type 6-aminohexanoate oligomer hydrolase"  
 /protein\_id="BAA05088.1"  
 /db\_xref="GI:488341"  
 /translation="MNTTFVHALTDIDGGIADVDPAPRLAGPPVGGGNDADFLLPV  
 STGREMLRDFDPFGVSGIAHYEEGPTGATVIHPAGARTAVDARGVAGLSGGYDFNH  
 AICLAGGAGYGLEAGAGVSDALLERLHRTGATVIELQLVSSAVIYDFARSATVYDPA  
 LGRALRPAVPEEPQGRAGAGMSACKVDMRTEITGQGAFFRGLDGVILAVVVP  
 NPVGIVDRACTVVRGNYDACTGYRRHPVDYQBAFAEQVPPVTEAGNTTISATVTV  
 RMPVELNQFAKQVHSSMRGIIQPFHTMDGDTLEAVTTDEIDLPTTPGSSRGRLSVN  
 ATALGAIASVWMDAVLEAGK"  
 complement(5674..5678)  
 /note="Shine-Dalgarno sequence"  
 14305..15190  
 /standard\_name="IS6100"  
 /note="RS-IB repeat sequence; insertion sequence"  
 /citation=[6]  
 /label=RS-IB  
 15191..15604  
 /note="RS-IB region; direct repeat from right end of  
 insertion sequence 6100 (IS6100)"  
 /rpt\_type=direct  
 15858..15861  
 /note="Shine-Dalgarno sequence"  
 15869..17047  
 /gene="nylB"  
 15869..17047  
 /gene="nylB"  
 /standard\_name="EII'"  
 /EC\_number="3.5.1.46"  
 /note="RS-IB repeated sequence; Sequence homologous to  
 RS-IA region"  
 /citation=[2]  
 /codon\_start=1  
 /transl\_table=11  
 /label=RS-IB  
 /product="6-aminohexanoate-dimer hydrolase"  
 /protein\_id="BAA05089.1"  
 /db\_xref="GI:488342"  
 /translation="MNTPTTGSHPARYPAAAGEPTLDSWQEPHNWFAHLGEMV  
 SAAVSRPVNAPCHALRIGATAAOLPDLEQRLTEOTYDAFLVLRGEVVAEYVRAGF  
 APDRHLLMSVSKSGCTVYVGLVDEGRIDPAQVTEVPELAGSVLDQVLDLDM  
 QISIDYNEDYVDFASEVQTHDRSAGWTRRGDPADTYEELTTLRGDGGTGEYCSEA  
 NTVLAWLIVERTVGLRYVEALSTYLMKADLDRDATITVDQTGFGFRANGVYSTARDL  
 ARVGRMLDGGVAGPGRVVSQGWVESYLAGGSRDMDTEGTFSAFPEGSTYRWMCCTG  
 NERGNVSGIGIHQNLWLDPRDTSVIVKLSSWPPDTRHWHGLQSGILLDVSRAILDV

17849..18728  
 /standard\_name="IS6100"  
 /note="RS-IC repeat sequence; insertion sequence"  
 /citation=[6]  
 /label=RS-IC  
 complement(29155..30034)  
 /standard\_name="IS6100"  
 /note="RS-ID repeat sequence; insertion sequence"  
 /citation=[6]  
 /label=RS-ID

misc\_feature  
 misc\_feature







The length in codons is given for each CDS. Usually the highest scoring similarity by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nihb.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid C88 Lies between and overlaps with cosmids C123 and C80 on the AseI-c genomic restriction fragment.

51101  
50000

```
FEATURES
source
gene
misc_f
CDS
```

## ORGANISM

```

/gene=SCC88.01"
note=Protein sequence is in conflict with the conceptual
translation: SCC88.01 possible glycosyltransferase, len:
>135 aa; N-terminal region similar to TR09XC92
(EMBL:AF14532) Klebsiella pneumoniae glycosyltransferase,
329 aa; fasta scores: opt: 214 z-score: 245.4 E(): 3e-06;
21.3% identity in 263 aa overlap and C-terminal region
similar to SWTAG6_BACSU (EMBL:X15200) Bacillus subtilis
telichoic acid biosynthesis protein F TagF, 746 aa; fasta
scores: opt: 1056 z-score: 1203.0 E(): 0; 41.5% identity
in 381 aa overlap"

```

## AUTHORS

```
/transfer_care=11
/product="putative glycosyltransferase"
```

```
/usr/local/bin/openssl s_client -connect 192.168.1.100:443 -sni_hostname 192.168.1.100
/translation="MRKENAGLAARTGMTRLAPESEYVAFVSDDDMIPDPAYRLMV
ESLDESGDFVTGNVQHINSTKIWQSPMHRILAGGAVRTHISRKKLLTDRIACNKV
EBDCEKJUDJCEBQHVKEVQVNTIACQENAVDTTCEMVNDIPECEAAGTIO
```

## JOURNAL

RRTEPKAVRDRAAAIVESVSRFFASRPGESGAELKHAYDHTALTGDLRIFLNVLPGDGA  
EFRAEFLRVINKYLDQDVPVVMELPATARIKWLLVRKHALDDLVDLIKAERRGDAVK  
TSGIRKVKYVSPTVEAAAAGLKKARRINPOIRIHAPLODTRWDGGRIRISGHAWVDO

## JOURNAL

LKLRGQWEEGTWRVIGIATAGLLRRRAVQVKPTRANHPPYQWLDDPRLLTPTPV  
NGALRTIVERVALLTGHRAVGADIEIDGEIRYP/LAGESVSRVTNKKSGERLHPV  
TLSDGAHTPTVTPVRPMRDVALLEPHOVDKEENAAAGGAPAAASRRNNSTQLVAVA  
PDGGRERTFVRFGIADCEMRILPASTAFNADNETALVSHNGAYLKSGRLOAKIT  
DGLHGGVETVTRVGVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV

+

LHPCSDLSLSEBVGQRRLKRDVYACGRTPQLDAVLFI SYNGKQY SDSHPHMERRE  
LRACADILEYLVLRGVDGDDTDPAQKQFWSEWFEALACRQIVTNHAPLPHMWERRE  
GOVIVOTHCMTMLKIGLIDIEAPKEDPOYHERLQOENRSTVLSNRRSTPILTKRAM  
GDKGIIETGPNKYLDAPDLASAREKEKRLGVSPDKRVLVLPATWRDDLSHRGQ  
FKFDURDVEDARLGDGDDHLLIRRHNSVDISIPAGNGFVHDVSEYDIDADLYLAA  
DLVYDYSVMSEYAHKLRKPMFLFFYDUEYRDKLGRVGFYFDEENDAPGPIRSTSEELY  
SALGDDKSVSAEYADRYRRFQELFCDDLDGGAARAVVDVIMLMEQAREV "  
3416 .3460  
3580 .5310

```
/gene="SCC88.02"
3580. 5310
/notes="SCC88.02"
similar to TR:CA651964 (EMBL:AL109661) Streptomyces
coelicolor hypothetical 100.7 kD protein SC6E10.10, 943
aa; fasta scores: 522 z-score: 556.2 E(): 1.4e-23;
31.8% identity in 471 aa overlap. Contains possible
N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB75372.1"
/db_xref="GI:6983733"
/translation="MTFSRLRTRSLTCTALAAAGMLVTTTAVAAADKPSGIE
VTPKSLPLDRQPERKQAPRMRAATAPRFDVGDGFSMLRLQDGSQVGLT
TAGDFTYITNVDDYERKQDVTTPGDDGNGGPEILTLSSSTGLSLFQSGWADSTGY
ATWSGTQYINQYKVPAGDVTGDRPDLARTYGGDLYLQGTGSDSPRGRVKVGG
GWEMFDQVGMNGDGIQDVVARTLGLYFYAGKGSATPLATRDLLGEGMTGN
QIIAADQVGLDGLGRTRSGDLYYRADTGNFEAKQOAGTGPQVFFEGAGTGN
PHFCKASLVNRYATLDESATGKYNLFLPGDLDNDCKGDLARDTSGLYDGLVAD
LLEYDGLVNYATLDESATGKYNLFLPGDLDNDCKGDLARDTSGLYDGLVAD
KGASLSPLKVGCGWNAVDKLVAGADFTGDRTDIVARTPAGKLYLRYGTGVSNSP
FAGKVDLGSQWQYNKLASPEDIDGGRADLLAVNSGGDLYRTSTYDGRFARVILG
TGMWYRNLY"
5327. 5362
complement(5387..8215)
/gene="SCC88.03c"
complement(5387..8215)
/gene="SCC88.03c"
/notes="SCC88.03c, possible transferase, len: 942 aa;
N-terminal region similar to hypothetical proteins, e. g.
TR:069853 (EMBL:AL023702) Streptomyces coelicolor
hypothetical 66.0 kD protein SC1C3.11, 586 aa; fasta
scores: opt: 1214 z-score: 1395.6 E(): 0; 40.7% identity
in 573 aa overlap and C-terminal region similar to
TR:069854 (EMBL:AL023702) Streptomyces coelicolor putative
transferase SC1C3.12, 697 aa; fasta scores: opt: 658
z-score: 754.0 E(): 0; 30.1% identity in 574 aa overlap.
Contains Pfam match at N-terminal region to entry PF00534
Glycosyl_transf.1, Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/product="putative transferase"
/protein_id="CAB75373.1"
/db_xref="GI:6983734"
/translation="MKITFLTWGDMGTEQAVYVTOATHLAPRHDIEVLVFKTRQ
PFVSDEVSVRYLVDRGKTERPVRESDLTAEDCRHLAGLPSDLISPKWEATFDALS
DVEQMRALRTIDTVLITSPALMSAVADLAAPSRTITQOEHRPSQLRGTEGPELLLR
LPAIDALVLTERTKOWLESGLKRAAPRLAIPNAIPEGFRPSRLSGTKTIVMPRLV
FDKQVDHAIQAFAPALPDHGRURIFGDPQMSRLRLNLIQGLHSDSVLELLGPSQHM
TEEMARASITILPSQDGEAFPLVLEAFAGVPVAVYDIVTGPAEIIHGEDGLLVPP
NDVESLAERISRLMGDEALLRSYGEKAHEGSTFAADYIVKQWEELETFVSRDDPR
RMAERADIAHRVAGGAGRFHAAVADRTAPSSDQORAREWIGASDRSLVRAGRL
SFVRDLOGSEIVQNFETVVEALESIGIPVLLDRDDNRRRLAVDAEQTVRKA
LAGATEKAVTAELKLPRTIAPGVLLAERLEAVEGAVGLRFPVPTSTRFLRGPAY
GCDIEFWQVEEGGQGVQVAPRSVAGPKLPLESTPDARTVRKDRYPTLEPLTRK
LSDITFFVDVAVTWDDSDPRQWRARRAALGLEAESDGEAAFRNRDELRYSL
RSLMAFAPWIRKIYLVTDQTPWLNTEHEGTEVYSHRDIFTDQDCLPTFNSHIESQ
LHHIDGLSEQFLVNDVFIQRPVGAORFFLPNGASREFWSPPTVPVGEETDEEYF
AAAKNNRALLERFGVTVANSFVHAPHLRRSVLESIEEDFPESVARTAAATPMRGWQD
ISMWSLHHHYGLFGASVPSIRCAIYIDVGTYSRHPELTLLAMRHGHDVFCLESQD
AEVPEHQARIVAEAFRLAFYFPVKSPYER"
complement(7124..7570)
/gene="SCC88.03c"
/notes="SCC88.03c"
Glycosyl transferases group 1, score 99.70, E-value
2.1e-27"
complement(8222..9211)
/gene="SCC88.04c"
complement(8222..9211)
/gene="SCC88.04c"
/notes="SCC88.04c, unknown, len: 329 aa"
```

```
gene
Query Match 3.8%; Score 52.4; DB 1; Length 42816;
Best Local Similarity 45.3%; Pred. No. 10;
Matches 191; Conservative 0; Mismatches 231; Indels 0; Gaps 0;
QY 942 ggtgatgaatgacctgccccagctgccccctctctaccaccactgaggtgacctacagcac 1001
Db 32480 GGTGATGAGGCCAAGACCGCGCCGTAGTCGCCCATGGCTCCAACCCCGACTACGACCC 32421
QY 1002 ggcgcacatcttcgaccgcatcaagcgcaagagcttcctcgtagaagacgccagcgggcc 1061
Db 32420 CAACGCCCTGGTGGCGGCATCTCCGCCAAGGACTACAGAAGCTCACCGCCGAAGGACTC 32361
QY 1062 caaggagaagctggagatctacaagccactgccccggtactgcatcgctccatgctgtc 1121
Db 32360 CAACTACCCGCTGCTCAACCGCGCCACCCAGGCCAGCGCGCCGCTCCATCTTCAA 32301
QY 1122 cctggagagacacacgctggcggcacacagactgctgtacgggacacacatgcagctcat 1181
Db 32300 GGTGCTCTCTCGCGCGGAGCGGTCGAGGCGCGGTACGACTTCGACGGCAACTACGACTG 32241
QY 1182 caccaggggcaaggggcgcgacgcacccacacctcatcagcaccgagttctcccgcgagct 1241
Db 32240 CTCAGGCGCTTACAACGTCGCGCGCCAGGCTTCAAGAAGCTTCAGAGTCGCCAACGAGGG 32181
QY 1242 ccactacaagtgagctgctctgccccctgattatctgcaagggtgactggagcaggtataa 1301
Db 32180 CATGATCACCATCGCGCGCGCTTGGAGCTCTCTCTGGGACACCGCTCTTCTACCGCTGTC 32121
QY 1302 cgaggcccgccctcccaacacgacagagagtcacagagagccctcgagcagaggacta 1361
Db 32120 CCACGAGGTGGAGAGGACGCGGGGATGAACCCCAAGAACCCCGCGACTACTTCTA 32061
QY 1362 ca 1363
Db 32060 CA 32059
```

Search completed: April 25, 2002, 12:13:26  
Job time: 7535 sec

Sequence	Strd Orig	zScore	EScore	Len	Documentation	..
gb_pat:AX351002	+	2513.00	3020.23	1.4e-107	1395	AX351002 Sequence 1 from Patent
gb_pat:AX351006	+	1567.00	2071.15	5.3e-107	936	AX351006 Sequence 5 from Patent
gb_pr:AL133463	+	954.00	1215.29	2.5e-59	84122	AL133463 Human DNA sequence
gb_pr:AK056709	+	930.50	1216.57	2.1e-59	3052	AK056709 Homo sapiens CDNA FLA
gb_pat:AX136231	+	929.50	1215.47	2.4e-59	2981	AX136231 Sequence 153 from Pat
gb_pat:AX351004	+	900.00	1193.94	3.8e-58	495	AX351004 Sequence 3 from Patent
gb_pr:BC017997	+	828.00	1087.01	3.5e-52	1617	BC017997 Homo sapiens, clone I
gb_pr:AF111168	-	634.00	781.56	3.6e-35	231464	AF111168 Homo sapiens serine
gb_pr:HSJ107712	+	321.00	374.53	1.7e-12	107057	AF1050320 Human DNA sequence
gb_sts:HS107712S	-	292.00	388.11	2.9e-13	523	AL1010084 H.sapiens STS from ge
gb_ro:MUS7SP1A	+	145.50	173.40	0.2687	4339	MB17276 Mouse thrombospondin 1
gb_htg:AL626774	+	144.50	134.65	38.69	199982	AL626774 Mus musculus chromo
gb_pr:AB005298	+	143.00	167.95	0.3403	5399	AB005298 Homo sapiens BAI 2 mR
gb_pat:EB21623	+	143.00	167.93	0.3419	5412	E21623 Human BAI gene and uti
gb_pat:EB21622	+	142.00	166.39	0.6606	5535	E21622 Human BAI gene and uti
gb_pr:AB005297	+	142.00	166.39	0.6606	5535	AB005297 Homo sapiens BAI 1 mR
gb_in:AF1060990	+	137.50	171.48	0.3438	1786	AF1060990 Drosophila melanogast
gb_in:AF135119	+	137.50	168.32	0.5151	2466	AF135119 Drosophila melanogast
gb_pr:AB005299	+	137.50	161.00	1.32	5218	AB005299 Homo sapiens BAI 3 mR
gb_pat:EB21624	+	137.50	160.91	1.33	5267	E21624 Human BAI gene and uti
gb_pr:AB011122	+	137.50	160.05	1.49	5749	AB011122 Homo sapiens mRNa fo
gb_ro:BTCPINTEGR	+	137.00	161.61	1.22	4580	X96389 B.taurus mRNa for proc
gb_ov:ESTLSTAGE	+	135.00	161.53	1.23	3522	L04278 xenopus laevis mRNa seq
gb_cm:BTAL33488	+	135.00	158.03	1.93	5039	AJ134488 Bos taurus mRNa for S
gb_cm:BTAL32106	+	133.50	164.81	0.8082	2054	AJ132106 Bos taurus mRNa for P
gb_pat:AX319857	+	133.50	154.66	2.97	5808	AX319857 Sequence 21 from Paten
gb_in:AC005717	-	133.50	133.37	45.56	51289	AC005717 Drosophila melanogast
gb_htg:AC0020007	+	133.50	129.83	71.79	73721	AC020007 Drosophila melanogast
gb_in:AE009032	+	133.50	122.07	194.15	163072	AE009032 Drosophila melanogast
gb_in:AE003804	+	133.50	117.52	348.03	259817	AE003804 Drosophila melanogast
gb_pat:AX211284	+	132.00	158.67	1.78	3144	AX211284 Sequence 1 from Patent
gb_pat:AX319856	+	132.00	157.45	2.08	3561	AX319856 Sequence 20 from Patent
gb_pat:AX342634	+	132.00	157.43	2.08	3567	AX342634 Sequence 31 from Patent
gb_pr:AF358666	+	132.00	157.15	2.16	3672	AF358666 Homo sapiens a disint
gb_in:CFWA42213	+	131.50	162.51	1.09	1982	U42213 Cryptosporidium wrairi
gb_ro:MMWPPER	+	131.00	165.54	0.7358	1358	X12905 Mouse mRNa for propedi
gb_pr:HSJ3125	+	130.00	154.60	3.00	3636	AK003125 Homo sapiens mRNa for
gb_pat:AR007441	+	130.00	153.59	3.41	4029	AR007441 Sequence 9 from patent
gb_pat:E32472	+	130.00	153.59	3.41	4029	E32472 Novel vascular smooth m
gb_ro:RATESA	+	130.00	153.59	3.41	4029	M8469 Rattusnorvegicus f-spo

```
351 TATCAATGGCGAGRTWCCAAATATCCAGGTCCACCATAGAGTGTGTCGAGC 400
134 lyProAspSerGluAlaAspLysaspGlnHisProGluAsnLysProSer 150
401 GTCTGTGACTCTGAAGCAGATAAAGATCAGCATCCGAGAGTAAGCCCGAGC 450
151 TrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSerLeuSe 167
451 TGGTCAGTCCCATCCCGCAGCTGGCGGGCCCTGGTGGCAGAGTCCCTGTC 500
167 rLeuAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrs 184
501 CTTGGCCAGGGCAACACAGCGGGAGCCAGGACTACRAGTACGACAGTACT 550
184 erAspAspSerAsnPhelLeuAsnProProArgGlyTrpAspHisThraLa 200
551 CAGACGACAGCAACTCTCTCAACCCCGCCAGGGGTGGGACCATACAGCC 600
201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217
601 CCAGGCCACCGGACTTTTGAACCAACAGATCAGCCAGATATGATTCAC 650
217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234
651 AGATGGCGAGGGTGACTGGAGTCTCTGGTCTGTCTGCAGCGCTCACCTGC 700
234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
701 GGAAGGCCAACCGAAGACGGACCCGGTCTTGGGTACGGTGCCTGCA 750
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluAspTh 267
751 ACAGAATCGAGGACCTGTGACCGTCCAAACTGCCAGGAATTTGAAGACAC 800
267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGlup 284
801 TTTTAGGACAGCTGCCACCGAAGTGGTCTGCTTCGGGAACGAGGAGT 850
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrp 300
851 TTAATGCCACCAACTGTTTGAAGTTGACACAGACAGCTGTGAGCGCTGG 900
301 MetSerCysLysSerGluPheLeuLysLysTyrMetHisLysValMetas 317
901 ATGAGCTGCAAAAGCGAGTCTTAAGAAGATACATGCACAGGTGATGAA 950
317 nAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSerT 334
951 TGACCTGCCAGCTGCCCTGCTCTACCCACTGAGGTGGCTACAGCA 1000
334 hrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
1001 CGGCGGACATCTTCGACCGCATCAAGCGCAAGGACTTCCGCTGGAGGAC 1050
351 AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367
1051 GCCAGCGGGCCCCAAGGAGAGCTGGAGATCTACAGCCCACTGCCCGGTA 1100
367 rCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGlnH 384
1101 CTGCATCCGCTCCATGCTGTCCCTGGAGAGCACCCAGCTGGCGGCACAGC 1150
384 isCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAla 400
1151 ACTGCTGCTACGGCGACACATCAGCTCATCACAGGGGGAGGGGGCG 1200
401 GlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyrLy 417
1201 GGCACGCCCAACTCATCAGCACCGAGTCTCCGCGGAGCTCCACTACAA 1250
417 sValAspValLeuProTrpIleLeCysLysGlyAspTrpSerArgTyrA 434
1251 GGTGGAGCTCTGCCCTGGATTATCTGCAAGGGTGACTGGAGCAGGTATA 1300
```

```
434 snGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSerProSer 450
1301 ACCAGGCGCGCCTCCCAACACGACAGAAAGTGCACAGAGAGCCCTCG 1350
451 AspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464
1351 GAGGAGACTACATCAAGCAGTTCCAGAGGCCAGGGAATAT 1392
seq_name: gb_pat:AX351006
seq_documentation_block:
LOCUS AX351006 936 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 5 from Patent WO0190179.
ACCESSION AX351006
VERSION AX351006.1 GI:18616374
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., Abuin,A.,
Zambrowicz,B. and Sands,A.T.
TITLE Novel human thrombospondin-like proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0190179-A 5 29-NOV-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source
1..936
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 231 a 283 c 248 g 171 t 3 others
ORIGIN
alignment_scores:
Quality: 1567.00 Length: 297
Ratio: 5.366 Gaps: 0
Percent Similarity: 98.316 Percent Identity: 97.980
alignment_block:
US-09-863-824-2 x AX351006 ..
Align seg 1/1 to: AX351006 from: 1 to: 936
1 MetValArgLeuAlaAlaGluLeuLeuLeuLeuGlyLeuLeuLeuLeuLeuLe 17
1 ATGTTGCGCCTGGCGCGGAGCTGCTGCTGCTGGGGCTGCTGCTGCT 50
17 uThrLeuHisIleThrValLeuArgGlySerGlyAlaAlaAspGlyProA 34
51 CACGCTGCACATCACCGTGCTGGCGGGCTCGGAGCCGCCGACGGGCCG 100
34 spAlaAlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnLeuAsn 50
101 ACGCGCGCGCGGCAACGCCAGCCAGCCAGCTGCAGAAATAACTCAAC 150
51 ValGlySerAspThrThrSerGluThrSerPheSerLeuSerLysGluAl 67
151 GTGGGAAGTGACACCATCAGAAACCCAGCTTTCTCTCTCCAAAGAAGC 200
67 aProArgGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgp 84
201 ACCAAGGAGCATCTGGACCACCGCTGCACACCAACCTTCCCCAGAC 250
84 roArgPheArgGlnGluThrGlyHisProSerLeuGlnArgAspPhePro 100
251 CGCGATTCGACACAGAGACGGGGCACCTTCATTGCAAAAGAGATTCCCC 300
101 ArgSerPheLeuLeuAspLeuProAsnPheProAspLeuSerLysAlaAs 117
301 AGATCCTTTTCTCTTGATCTACCAAACTTTTCCAGATCTTTTCCAAAGCTGA 350
```

```

117 pilleAsnGlyGln***ProAsnIleGlnValThrIleGluValValAspG 134
|||||
351 TATCAATGGCGAGRTWCATAATATCCAGGTACCATAGAGGTGCTCGACG 400
|||||
134 lyProAspSerGluAlaAspLysAspGlnHisProGluAsnLysProSer 150
|||||
401 GTCTGAGCTCTGAAGCAGATAAAGATCAGCATCCGGAGAAATAGCCGAG 450
|||||
151 TrpSerValProSerProAspTirpArgAlaTrpTrpGlnArgSerLeuSe 167
|||||
451 TGGTCAGTCCCATCCCCGAGCTGGGGCCCTGGTGGCAGAGGTCCCTGTC 500
|||||
167 rLeuAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrS 184
|||||
501 CTTGGCCAGGCGCAACAGCGGGACAGGACTACRAGTACGACAGTACCT 550
|||||
184 erAspAspSerAsnPhelLeuAsnProProArgGlyTyrAspHisThrAla 200
|||||
551 CAGACGACAGCAACTTCTCAACCCCGCCAGGGGTGGGACCATACAGCC 600
|||||
201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217
|||||
601 CCAGCCACCGGAGCTTTTGAACCAACAGATCAGCCAGAAATGATGATCC 650
|||||
217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234
|||||
651 AGATGGCGAGGTGACTGGAGTCTCTGGTCTGCTGCAGCGTCACTCCG 700
|||||
234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
|||||
701 GGAACGGCAACAGAAACGGACCGCGTCTTGGCTAGCGTGCACGTCA 750
|||||
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluAspTh 267
|||||
751 ACAGATCGGAGGACTGTGACCGTCCAACTGCCAGGAATGAAGACAC 800
|||||
267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluP 284
|||||
801 TTTTAGGACAGCTGCCACCGAAGTGAGTCTGCTTGGCGGAAGCGAGGT 850
|||||
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCys 297
|||||
851 TTAATGCCACCAAACTTTTGAAGTGTGCTGCCAGCATGT 891
|||||

```

seq\_name: gb\_pr:AL133463

seq\_documentation\_block:  
LOCUS AL133463 84122 bp DNA linear PRI 15-MAR-2001  
DEFINITION Human DNA sequence from clone RP11-149118 on chromosome 20.  
Contains the 3' end of a novel gene, ESTs, STSS and GSs, complete sequence.

ACCESSION AL133463  
VERSION AL133463.16 GI:10443352  
KEYWORDS HTG.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 84122)

REFERENCE

AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

COMMENT

On Oct 1, 2000 this sequence version replaced gi:10178396.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:  
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human 20  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20

IMPORTANT: This sequence is not the entire insert of clone  
RP11-149118 It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true right end of clone RP4-585114 is at 84023 in this sequence.  
The true left end of clone RP5-107712 is at 100 in this sequence.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e. phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP11-149118 is from  
the library RPCI-11.1 constructed by the group of Pieter de Jong.  
For further details see  
http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6.  
Location/Qualifiers

source

```

1. .84122
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP11-149118"
/clone_lib="RPCI-11.1"
/join("<1.159,9050.9314,17956.18099,21790.21879,
28358..30067")
/gene="bAl49118.1"
/note="match: ESTs: Em:AL191256 Em:R44763 Em:W86073
Em:AI082824 Em:AI886394 Em:R52250 Em:AV708785 Em:W07772
Em:AV660737 Em:BF302850 Em:AW426875 Em:W86257 Em:BF302850
Em:AW873508 Em:BE287313 Em:AW495535 Em:AW435648
Em:AW435639"
/product="bAl49118.1 (novel protein)"
/evidence="not_experimental"
1. .30067
/gene="bAl49118.1"
/join("<1.159,9050.9314,17956.18099,21790.21879,
28358..28875")
/gene="bAl49118.1"
/note="continued from dJ107712.1 in Em:AL050320
match: proteins: tr:O95432 tr:Q9RTS5"
/evidence="not_experimental"
/product="bAl49118.1 (novel protein)"
/db_xref="GI:13374941"
/translation="HQAAGPPRRFRQETGHPSLQRDPFRSFLDLNFPDLSKAD
INGQNPDIQVTEIIVNDPDSQADKQHPENKPSWSPDWRAMWGRSLARNSGD
QDYKYDSTDSDSNFNPPRGWDHATGHRTFETKQDPEYDSTDGEGSLWSVCSYTC
GNGNOKRTRSCGACTSRTCDRNPCEIEDFTTAATEVSLLAGSEFNATKLF
VPTDCSERMNSCKSEFLKKYMKHVMNDLPSCSCSYPTVEAYSTADLFDRIKKRDFRK
DASGPKLEIYKPTARYCIRSMISLESTLAAHCCYGDNMOLITRGKAGCTPNLIS
TEFSALHYKVDVLPWICKGDWSRYNEARPPNNGOKCTESDSEDIYIKQFQAREY"
complement(1..81)
/note="match: STS: Em:HS107712S"
437.504
/note="34 copies 2 mer aa 66% conserved"
445.510
/note="2 copies 33 mer 93% conserved"
885.1046
/note="MIR repeat: matches 35.189 of consensus"
1202.1494
/note="AluX repeat: matches 1.291 of consensus"
2642.2951
/note="AluJo repeat: matches 1.311 of consensus"
3460.3507

```

mRNA

gene

CDS

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```
/note="24 copies 2 mer tg 79% conserved"
4065..4106
/repeat_region
/note="21 copies 2 mer cc 76% conserved"
complement(4442..5302)
/misc_feature
/note="match: GSS: Em:AQ741486"
4908..5154
/repeat_region
/note="MIR repeat: matches 34..238 of consensus"
5180..5672
/misc_feature
/gene="ba1491i18.1"
/note="match: GSS: Em:AQ755240"
7941..8067
/repeat_region
/note="MIR repeat: matches 1..140 of consensus"
8078..8510
/repeat_region
/note="MIR repeat: matches 81..519 of consensus"
8538..8624
/repeat_region
/note="MIR repeat: matches 170..262 of consensus"
complement(8621..9044)
/misc_feature
/note="match: GSS: Em:AQ182554"
complement(8758..9044)
/misc_feature
/note="match: GSS: Em:AQ133736"
9887..10251
/repeat_region
/note="MIR repeat: matches 19..371 of consensus"
10536..10673
/repeat_region
/note="L2 repeat: matches 2563..2708 of consensus"
10826..11052
/repeat_region
/note="MIR repeat: matches 2..261 of consensus"
11277..11343
/repeat_region
/note="MER5A repeat: matches 99..166 of consensus"
11344..11517
/repeat_region
/note="MER5A repeat: matches 1..182 of consensus"
11520..11698
/repeat_region
/note="MER5B repeat: matches 2..178 of consensus"
11773..11825
/repeat_region
/note="MER5A repeat: matches 136..187 of consensus"
12156..12212
/repeat_region
/note="MER5A repeat: matches 132..189 of consensus"
12248..12291
/repeat_region
/note="MER5A repeat: matches 9..52 of consensus"
12443..12595
/repeat_region
/note="AluTo/FRAM repeat: matches 162..310 of consensus"
complement(13043..13439)
/misc_feature
/note="match: GSS: Em:B33790"
13397..13552
/repeat_region
/note="MIR repeat: matches 35..193 of consensus"
14382..14420
/repeat_region
/note="MLTII repeat: matches 514..547 of consensus"
14421..14739
/repeat_region
/note="AluTo repeat: matches 1..310 of consensus"
14740..15085
/repeat_region
/note="MLTII repeat: matches 117..514 of consensus"
16534..16892
/repeat_region
/note="THE1B repeat: matches 1..364 of consensus"
16900..17163
/repeat_region
/note="L2 repeat: matches 2491..2747 of consensus"
17207..17360
/repeat_region
/note="MLTII repeat: matches 415..568 of consensus"
18198..18489
/repeat_region
/note="AluSq repeat: matches 13..301 of consensus"
18940..19047
/repeat_region
/note="MER5C repeat: matches 1..132 of consensus"
19018..19116
/repeat_region
/note="3 copies 33 mer 82% conserved"
19103..19197
/repeat_region
/note="MER5C repeat: matches 119..89 of consensus"
19785..20098
/repeat_region
/note="AluY repeat: matches 2..306 of consensus"
20136..20382
/repeat_region
/note="MIR repeat: matches 8..261 of consensus"
21260..21460
/repeat_region
/note="MIR repeat: matches 2..219 of consensus"
22331..22644
/repeat_region
/note="AluY repeat: matches 1..302 of consensus"
23110..23419
```

```
/note="AluSq repeat: matches 3..312 of consensus"
2396..24102
/note="LIR33 repeat: matches 415..521 of consensus"
24208..24360
/repeat_region
/note="MIR repeat: matches 38..192 of consensus"
24986..25364
/note="LIR16B repeat: matches 1..449 of consensus"
25661..25780
/note="LIR16C repeat: matches 1..134 of consensus"
25792..26153
/note="MER92B repeat: matches 272..631 of consensus"
26154..26431
/note="AluSq repeat: matches 1..278 of consensus"
26432..26701
/note="MER92B repeat: matches 2..272 of consensus"
26702..26928
/note="LIR16C repeat: matches 150..387 of consensus"
27503..27831
/note="AluJb repeat: matches 1..312 of consensus"
27890..27960
/note="L1MC4 repeat: matches 7716..7789 of consensus"
27988..28097
/note="L1MC4 repeat: matches 7857..7977 of consensus"
complement(29592..30009)
/misc_feature
/note="match: STS: Em:G21248"
29986..29991
/polyA_signal
/gene="ba1491i18.1"
30009
/polyA_site
/gene="ba1491i18.1"

alignment_scores:
  Quality: 954.00      Length: 204
  Ratio: 5.074        Gaps: 2
  Percent Similarity: 92.157  Percent Identity: 87.745

alignment_block:
US-09-863-824-2 x AL133463

Align seg 1/1 to: AL133463 from: 1 to: 84122

262  ProGlylleGluAspThrPheArgThrAlaAlaThrGluValSerLeuLe 278
      |||||:::
28276  CCTGGTGC.....ACATGCTCTTGAAGCACCTCTCTGTG 28310

278  uAlaGlySerGluGluPheAsnAlaThrLysLeuPheGluVal....AspT 294
      |||||:::
28311  GGCCGGTTGTACACGCGGTGAGAGGGTTCTCTTTGGCCTTTTCCAGACA 28360

294  hrAspSerCysGluArgTrpMetSerCysLysSerGluPheLeuLysLys 310
      |||||:::
28361  CAGACAGCTGTGCGCTGGATGAGCTGCAAAAGCGAGTCTCTTAAGAAG 28410

311  TyrMetHisLysValMetAsnAspLeuProSerCysProCysSerTyrPr 327
      |||||:::
28411  TACATGCAACAAGGTGATGATGACCTGCCAGCTGCCCTCTCTCTACCC 28460

327  oThrGluValAlaTyrSerThrAlaAspIlePheAspArgIleLysArgL 344
      |||||:::
28461  CACTGAGTGGCGCTACAGCAGCGCCGACATCTTCGACCGCATCAAGCGCA 28510

344  ysAspPheArgTrpLysAspAlaSerGlyProLysGluLysLeuGluLe 360
      |||||:::
28511  AGGACTTCGCTGGAAGGACGCCAGCGGCCCAAGGAGAGCTGGAGATC 28560

361  TyrLysProThrAlaArgTyrCysIleArgSerMetLeuSerLeuGluSe 377
      |||||:::
28561  TACAAGCCCACTGCCGGTACTGTCATCCGCTCCATGCTGTCTCTGAGAG 28610

377  rThrThrLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMetGlnLeuI 394
      |||||:::
28611  CACCACGCTGGCGGCACAGCACTGCTGTACGGCGCAACATGCAGCTCA 28660

394  leThrArgGlyLysGlyAlaGlyThrProAsnLeuIleSerThrGluPhe 410
```

```

|||||
28661 TCACCAAGGGCAAGGGGGCGGCACGCCAACCTCATCAGCACCGAGTTC 28710
411 SerAlaGluLeuHisTyrIysValaspValLeuProTrrPrlleIleCysIy 427
28711 TCCGGGGAGCTCCACTACAAAGGTGGACCTCTCCCTCGATATTCGCAA 28760
427 sGlyAspTrrPrrSerArGtYrAsnGluAlaArgProProAsnAsnGlyGlnL 444
28761 GGGTCACTGGACAGGTATACAGAGCGCGGCTCCCAACACGACGACAGA 28810
444 yScyThrGluSerProSerAspGluAspTrrIleYlsGlnPheGlnGlu 460
28811 AGTGCACAGAGAGCGCCCTCGGACGAGGACTACATCAAGCAGTTCCAAGAG 28860
461 AlaArgGluTyr 464
28861 GCCAGGGAATAT 28872

```

seq\_name: gb\_pr:AK056709

seq\_documentation\_block:

LOCUS AK056709 3052 bp mRNA linear PRI 31-OCT-2001  
 DEFINITION Homo sapiens cDNA FLJ32147 fis, clone PLACE5000116.  
 ACCESSION AK056709  
 VERSION AK056709.1 GI:16552189  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens placenta cDNA to mRNA, clone\_lib:PLACES  
 clone:PLACES000116.

#### ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,  
 Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,  
 Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
 Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
 Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
 WagaTsuna,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,  
 Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and  
 Isogai,T.

#### TITLE

NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 3052)  
 Isogai,T., Otsuki,T. and Sugiyama,T.  
 Direct Submission  
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: RAB and  
 HRI.

#### FEATURES

source Location/Qualifiers  
 1..3052  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="PLACE5000116"  
 /tissue\_type="placenta"  
 /clone\_lib="PLACES"  
 /note="cloning vector: pME18SFL3"  
 622 a 946 c 910 g 574 t

#### BASE COUNT ORIGIN

alignment\_scores:  
 Quality: 930.50 Length: 500  
 Ratio: 2.926 Gaps: 18  
 Percent Similarity: 63.600 Percent Identity: 41.600

```

alignment_block:
US-09-863-824-2 x AK056709
..
Align seg 1/1 to: AK056709 from: 1 to: 3052

25 ArgGlySerGlyValAlaAspGlyProAspAlaAlaGlyAsnAlase 41
|||||
428 AGAGGTACTCCGTTGGCTGGAGCTCAGAAAGCTCGCGGATGGC... 475
|||||
41 rGlnAlaGlnLeuGlnAsnAsnLeuVal..... 51
|||||
476 .....CAACACACCTTGAGTACCCCTAACCCCTGATACCCAG 512
|||||
52 .....GlySerAspThr 55
|||||
513 GCTTCAGCCTCCCCAGATCCTAGGCCTCTGAGGGAGAGGAGGAGGCACG 562
|||||
56 ThrSerGluThrSerPheSerLeuSerLysGluAlaProArgGluHisLe 72
|||||
563 ACTGCTCCCCAGACCCACCTCGCAGGCAGAGCTACACCAACATGGATGT 612
|||||
72 uAspHisGlnAlaAlaHisGlnProPheProArgProArgPheArgGlnG 89
|||||
613 GGACTGTCTACTGAGCCAGCAGCCCTGACCCAGG.....G 647
|||||
89 lu.ThrGlyHisProSerLeuGlnArgaspPheProArgSerPheLeuLe 105
|||||
648 AATGCCAGCCCTCCAGGACCAGGAGGTACTCC.....TTGCTGCT 691
|||||
105 uAspLeuProAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln* 122
|||||
692 GGAGCTGCAGAGAGCTGCCAGAAATGGTCACGCAACCTTGATACCCCTA 741
|||||
122 **ProAsnIleGlnValThrIleGluValValaspGlyProAspSerGlu 138
|||||
742 ACCCTGATAAACAGTCACCATCAAGTGGTGGAGGAGCCGCCAGGCCGAG 791
|||||
139 AlaAspLysasp.....GlnHisProGluAsnLysPro..... 149
|||||
792 GTGTCGATAGACTCTGTGGCTGAGCCCAATCCCGCCGCCAGGATAC 841
|||||
150 ....SerTrpSerValProSerProAspTrpArgAlaTrpTrp..... 162
|||||
842 CCTTAGCTGG...CTGCCGCCCTC...TGCTCTCTCTGGGAGACT 885
|||||
163 .....GlnArgSerLeuSerLeuAlaArgAlaAsnSerGly..... 174
|||||
886 ACAAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935
|||||
175 .....AspGlnAspTyr..... 178
|||||
936 GAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 985
|||||
179 .....***TyrAspSerThrS 184
|||||
986 GGACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1035
|||||
184 erAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAla 200
|||||
1036 CAGAC.....AACTGGGACACAGGCTGG.....CTGGCC 1064
|||||
201 ProGlyHisArgThrPheGluThrLysaspGlnProGluTyrAspSerTh 217
|||||
1065 CCCGGGAGTGGGTCTTC.....AAGGATTCCTGCAGCTAGCAGC...TA 1105
|||||
217 rAspGlyGluGlyAspTrrPrrSerLeuTrpSerValCysSerValThrCysG 234
|||||
1106 TGAGCCTCAGAGAGGAGTGGAGTCCCTGGTCTCCCTGCAGTGGGAACGCA 1155
|||||
234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
|||||
1156 GCACCTGGCAAGCAGCAGAGGAGCTCGGCCCTGTGGCTATGGCTGCACCTGCC 1205
|||||

```





```

1673 GCCTCACGCTGTGCTCCTCCCAACAACGGCGGACCTGTCCACCGACAAC 1722
449 ProSerAspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464
III ::::::::::::::::::::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1723 CCCCTGGAGGAGGAGTACCTAGCACAGTTTGCAGGAGGCCAAGGAGTAC 1770

seq_name: gb_pat:AX351004

seq_documentation_block:
LOCUS AX351004 495 bp DNA linear
DEFINITION Sequence 3 from Patent WO0190179.
ACCESSION AX351004
VERSION AX351004.1 GI:18616373
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
REFERENCE 1 (sites)
AUTHORS Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., A.
Zambrowicz,B. and Sands,A.T.
TITLE Novel human thrombospondin-like proteins and polynuc
encoding the same
JOURNAL Patent: WO 0190179-A 3 29-NOV-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source
1. 495
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 126 a 156 c 139 g 74 t
ORIGIN

alignment_scores:
Quality: 900.00 Length: 164
Ratio: 5.488 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-863-824-2 x AX351004 ..

Align seg 1/1 to: AX351004 from: 1 to: 495

301 MetSerCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAs 317
|||||
1 ATGAGCTGCAAAAGCGAGTCTTAAAGAAAGTACATGCACAAGGTGATGAA 50
|||||
317 nAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSerT 334
|||||
51 TGACCTGCCAGCTGCCCTTGCTCTCTACCCCACTAGGTGGCCCTACAGCA 100
|||||
334 hrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
|||||
101 CGCGCGACATCTTCAGCCGCATCAACGGCAAGGACTTCCGCTGGAAGGAC 150
|||||
351 AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367
|||||
151 GCCAGGGGCCCAAGGAGAAGCTGGAGATCTACAAGCCCACTGCCCGGTA 200
|||||
367 rCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGlnH 384
|||||
201 CTGCATCGCTCCATGCTGCTCCCTGGAGAGCACACGCTGGCGGCACAGC 250
|||||
384 iCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAla 400
|||||
251 ACTGCTGTCTACGGCGCAACAATGCAGCTCATCACAGGGGCAAGGGGGCG 300
|||||
401 GlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyrLy 417
|||||
301 GGCAGCGCCCAACTCATCAGCACCGGAGTCTCTCCCGGAGCTCCACTCAA 350
|||||
417 sValAspValLeuProTrpIleIleCysLysGlyAspTrpSerArgTyrA 434
|||||

```

```
351 GGTGAGCTCCTGCTGATTTCTGCAAGGGTGACTGGAGCAGGTATA 400
434 snGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSerProSer 450
401 ACAGGCGCGGCTCCACACGACGACAGAGTGACACAGAGGCCCTCG 450
451 AspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464
451 GACGAGGACTATCATCAGCAGTTCCAGAGGCCAGGGAATAT 492
```

```
seq_name: gb_pr:BC017997
seq_documentation_block:
LOCUS BC017997 1617 bp mRNA linear PRI 06-DEC-2001
DEFINITION Homo sapiens, clone IMAGE:4252124, mRNA, partial cds.
ACCESSION BC017997
VERSION BC017997.1 GI:17389973
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1617)
Direct Submission
Strausberg, R.
Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 35 Row: f Column: 10  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Location/Qualifiers  
1..1617  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4252124"  
/tissue\_type="Skin, melanotic melanoma, high MDR."  
/clone\_lib="NIH\_MGC\_62"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
<1..457  
/codon\_start=2  
/product="Unknown (protein for IMAGE:4252124)"  
/protein\_id="AAH17997.1"  
/db\_xref="GI:17389974"  
/translation="KVMNDLPSCPSYFEVAYSTADIFDRKKDFRWKDSGPKEK  
LEIYRTYICIRMSLESTLAAHCCYGDNMOLITRGAGTGNLITFEFAELH  
YKVDLPVITCKGDSRYNEARPNNGKCTESPSDEYIKQFQAREY"

BASE COUNT 506 a 348 c 391 g 372 t  
ORIGIN

alignment\_scores:  
Quality: 828.00 Length: 151  
Ratio: 5.483 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-863-824-2 x BC017997
Align seg 1/1 to: BC017997 from: 1 to: 1617

314 LysValMetAsnAspLeuProSerCysProCysSerTyrProThrGluVa 330
|||||
2 AAGGTGATGAATGACCTGCGCAGCTGCTCCCTACCCCACTGAGGT 51
|||||
330 lAlaTyrSerThrAlaAspIlePheAspArgIleLysArgLysAspPheA 347
|||||
52 GGCTACAGCAGCGCCGACATCTTCGACCGCATCAAGCGCAAGGACTTCC 101
|||||
347 rGTpLysAspAlaSerGlyProLysGluLysLeuGluIleTyrLysPro 363
|||||
102 GCTGAAGGAGCGCCAGCGGCCCAAGGAGAAGCTGGAGATCTACAAGCCC 151
|||||
364 ThrAlaArgTyrCysIleArgSerMetLeuSerLeuGluSerThrLe 380
|||||
152 ACTGCCGGTACTGCATCCGCTCCATGCTGTCTCCCTGGAGAGCACCAGCT 201
|||||
380 uAlaAlaGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArg 397
|||||
202 GCGGCACAGCACCTGCTGCTACGGCGCACACATGCAGCTCATCACCAGG 251
|||||
397 lYsGlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGlu 413
|||||
252 GCAAGGGGGCGGCGACGCCCAACCTCATCAGCAGCGAGTTCTCCGCGGAG 301
|||||
414 LeuHisTyrLysValAspValLeuProThrIleLysGlyAspTr 430
|||||
302 CTCCACTACAAGGTGAGCGTCTGCCCTGGATTATCTCAAGGTGACTG 351
|||||
430 pSerArgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysCysThr 447
|||||
352 GAGCAGGTATACAGAGCGCGCGCTCCCAACACGACAGAGTGACAG 401
|||||
447 LuSerProSerAspGluAspTyrIleLysGlnPheGlnGluAlaArgGlu 463
|||||
402 AGAGCGCCCTCGGACGAGGACTACATCAAGCAGGTTCCTCAAGAGGCCAGGAA 451
|||||
464 Tyr 464
452 Tat 454
seq_name: gb_pr:AF111168
seq_documentation_block:
LOCUS AF111168 231464 bp DNA linear PRI 14-JUL-1999
DEFINITION Homo sapiens serine palmitoyl transferase, subunit II gene,
complete cds; and unknown genes.
ACCESSION AF111168
VERSION AF111168.2 GI:5468517
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 231464)
Multimegabase Sequencing Group.
Complete sequence of the gene for serine palmitoyltransferase,
subunit II found on human chromosome 14
Unpublished
2 (bases 1 to 231464)
UW Multimegabase Sequencing Group.
Direct Submission
Submitted (27-DEC-1998) Department of Molecular Biotechnology, Box
357730 University of Washington, Seattle, Washington 98195, USA
On Jul 14, 1999 this sequence version replaced gi:4186181.
Sequencing methodology: high redundancy shotgun with plasmids.
Interspersed Repeats were identified with RepeatMasker (available
from http://ftp.genome.washington.edu/RN/RepeatMasker.html).
```

FEATURES	Location/Qualifiers	
source	1. .231464	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="14"	
	/map="14q24.3-31"	
	/clone="BAC 50114"	
	/clone_lib="RPCI-11"	
	9. .186	
repeat_region	/rpt_family="LIMD1"	complement(9631. .9708)
repeat_region	200. .494	/rpt_family="LINE2"
repeat_region	/rpt_family="AluSx"	complement(10461. .10762)
repeat_region	500. .1283	/rpt_family="AluSx"
repeat_region	/rpt_family="AluSg"	complement(10804. .11077)
repeat_region	1284. .1581	/rpt_family="AluJb"
repeat_region	/rpt_family="AluSg"	complement(11156. .11306)
repeat_region	1582. .1774	/rpt_family="MIR"
repeat_region	/rpt_family="AluJb"	complement(11493. .11615)
repeat_region	1775. .2058	/rpt_family="FLAM_C"
repeat_region	2059. .2934	11800. .12161
repeat_region	/rpt_family="AluJb"	/rpt_family="THE1C"
repeat_region	/rpt_family="AluJb"	complement(12406. .12695)
repeat_region	/rpt_family="AluSg"	12746. .12920
repeat_region	complement(2937. .3255)	/rpt_family="BC200"
repeat_region	/rpt_family="MER4A"	12921. .13217
repeat_region	3323. .3687	/rpt_family="AluSx"
repeat_region	/rpt_family="L1"	13236. .13751
repeat_region	3688. .3977	/rpt_family="LIME1"
repeat_region	/rpt_family="AluSg"	13783. .14084
repeat_region	4009. .4314	/rpt_family="AluSx"
repeat_region	/rpt_family="AluJo"	14085. .14403
repeat_region	4321. .4934	/rpt_family="LIME2"
repeat_region	/rpt_family="L1"	14405. .14717
repeat_region	4940. .5236	/rpt_family="AluSx"
repeat_region	/rpt_family="AluSp"	14795. .14857
repeat_region	5238. .5349	/rpt_family="LIME2"
repeat_region	/rpt_family="LIME1"	complement(14884. .14905)
repeat_region	5350. .5650	/rpt_family="AT-rich"
repeat_region	5868. .5924	complement(15106. .15393)
repeat_region	/rpt_family="AluSg"	/rpt_family="AluSx"
repeat_region	5869. .5940	complement(15606. .15818)
repeat_region	/rpt_family="CAAAA)n"	complement(15835. .16131)
repeat_region	6086. .6271	/rpt_family="AluSc"
repeat_region	/rpt_family="AluJb"	complement(16775. .17073)
repeat_region	complement(6500. .6588)	/rpt_family="AluY"
repeat_region	/rpt_family="5S"	17491. .17777
repeat_region	complement(6715. .6815)	/rpt_family="AluJo"
repeat_region	/note="low quality data"	17778. .17804
repeat_region	complement(6914. .7458)	/rpt_family="(CA)n"
repeat_region	7523. .7707	complement(17815. .17882)
repeat_region	/rpt_family="LIMB6"	/rpt_family="LIMB2"
repeat_region	7708. .8007	17883. .18175
repeat_region	/rpt_family="AluSx"	/rpt_family="AluSx"
repeat_region	8233. .8305	18177. .18251
repeat_region	/rpt_family="(TA)n"	/rpt_family="(TAGA)n"
repeat_region	9186. .9230	complement(18244. .18306)
CDS	complement(join(9248. .9470,10060. .10240,16529. .16722,25330. .25484))	/rpt_family="LIMB1"
	/note="Intron-exon boundaries defined in relation to an EST contigs that includes AA397580, A1138314, AA399545, AA868184, all from testis. The closest match in BLASTX is to E. coli pyrroline-5-carboxylate reductase."	complement(18310. .18609)
	/codon_start=1	/rpt_family="AluSx"
	/product="unknown"	complement(18624. .18771)
	/protein_id="AAD09625.1"	/rpt_family="LIMB3"
	/db_xref="GI:4186186"	complement(18765. .18847)
	/translation="MDMLQDLESLOFEXGVPEEDRIWLYLOGRSGLMTEACAHATFFCKLLYNLRASLNKNOSSRHLSIGSLNSATPEEFKVIIGGGHGLKGLACTLLQLGPILASLSIRPPTLGLKLGKICFYHNADLVSWADVIFLCLLPOLPNICVEIYTSLEKASIVSFVAAPILPRUKLLNLNHTNLRPOQYDEDSVSWGANKGVAAALQDPTILQATCPSAGHGSTYFWHSFENYFPVGPLLSY"	/rpt_family="L1"
		18858. .19501
		/rpt_family="SVA"
		complement(19503. .20417)
		/rpt_family="L1"
		complement(20104. .20625)
		/rpt_family="L143_Send"
		complement(20724. .21029)
		/rpt_family="AluJb"
		complement(21030. .21122)
		/rpt_family="(TAAA)n"
		complement(21125. .21423)
		/rpt_family="AluJo"
		complement(21486. .21662)
		/rpt_family="L191_Send"
		complement(21664. .21966)
		/rpt_family="AluSg"
		complement(22042. .22342)

```
repeat_region      /rpt_family="AluY"
                    complement(22350..22937)
repeat_region      /rpt_family="L143_5end"
                    complement(22975..23090)
repeat_region      /rpt_family="FLAM_C"
                    complement(23164..23287)
repeat_region      /rpt_family="L143_5end"
                    complement(26429..26704)
repeat_region      /rpt_family="L1MB5"
                    complement(26766..27080)
repeat_region      /rpt_family="AluX"
                    complement(28232..28432)
repeat_region      /rpt_family="MER20"
                    complement(28528..28689)
repeat_region      /rpt_family="AluJb"
                    complement(28692..28984)
repeat_region      /rpt_family="AluX"
                    complement(29016..29108)
repeat_region      /rpt_family="MIR"
                    complement(join(30211..30231,30965..31069,31601..31690,
                    32276..32362,36437..36596,36857..36898,37854..37988,
                    38439..38514,40403..40476,43661..43688,45155..45257,
                    45714..45747,46844..46936,51089..51145,51888..51952,
                    53782..53928,55894..55996,56605..56697))
                    /note="Intron-exon boundaries defined by an EST contig
                    that includes W33552, W80307, AA780270, R53310,
                    and AA437679. There is a short form of this coding
                    sequence that terminates at position 188433, based on EST
                    AA780270. The closest similarity by BLASTX is to a C.
                    elegans hypothetical protein."
```

## alignment\_scores:

```
Quality: 634.00      Length: 172
Ratio: 4.227         Gaps: 0
Percent Similarity: 87.209      Percent Identity: 60.465
```

## alignment\_block:

US-09-863-824-2 x AF111168/rev ..

Align seg 1/1 to reverse of: AF111168 from: 1 to: 231464

```
293 AspThrAspSerCysGluArgTrpMetSerCysLysSerGluPheLeuIys 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78708 GATGTGGACAGCTGTGAGAAGTGGTGAACCTGCAAGACGACTTCTTAAT 78659

309 sLysTyrMetHisLysValMetAsnAspLeuProSerCysProCysSerT 326
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78658 CAAGTATCTGACGACGATGCTGGGAGCCTGCCAGCTGCCGCTGTGCCT 78609

326 yrProThrGluValAlaTyrSerThrAlaAspIlePheAspArgIleLys 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78608 ACCCACTGGAGGCCATGGACAGCCCTGTGAGCCTACAGACGAGCACCAG 78559

343 ArgLysAspPheArgTrpLysAspAlaSerGlyProLysGluLysLeuCl 359
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78558 GGCCGCGAGCTTCGGTGGAGGAGTGCACGTGGCCCTCGCGAGCGCTGGA 78509

359 uileTyrLysProThrAlaArgTyrCysIleArgSerMetLeuSerLeuG 376
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78508 CATCTACACGCCACCGCGCGCTTCTGCTGCTTCCATGCTGTCTGGGG 78459

376 luSerThrThrLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMetGln 392
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78458 AGAGCAGCACACTGGCGCCCGCAGCACTGCTGTATGACGAGGACGACCGG 78409

393 LeuIleThrArgGlyLysGlyAlaGlyThrProAsnLeuIleSerThrGl 409
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78408 CTGCTGACCGTGGCAAGGCGCGCGCATGCCCACTCATCAGCACCAG 78359

409 uPheSerAlaGluLeuHisTyrLysValAspValLeuProTrpIleIleC 426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78358 CTTCCTACCTAAGCTGCACCTCAAGTTCGACACGACGCGCCCTGGATCCTGT 78309
```

```
426 ysLysGlyAspTrpSerArgTyrAsnGluAlaArgProProAsnAsnGly 442
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78308 GCAAGGGGACTGGAGCGCTCCACGCTGTGCTCCCTCCCAACACGCG 78259

443 GlnLysCysThrGluSerProSerAspGluAspTyrIleLysGlnPheGl 459
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78258 CGAGCTGTCACCGACCAACCCCTGGAGGAGGAGTACCTAGCACAGTTGCA 78209

459 nGluAlaArgGluTyr 464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78208 GGAGGCCAAGGAGTAC 78193
```

seq\_name: gb\_pr:HSU107712

seq\_documentation\_block:

LOCUS HsU107712 107057 bp DNA linear PRI 04-APR-2001  
Human DNA sequence from clone RP5-107712 on chromosome 20. Contains  
part of a novel gene, ESTs, STSs, GSSs and a CpG island, complete  
sequence.

ACCESSION AL050320  
VERSION AL050320.19 GI:7263998  
KEYWORDS HTG: CpG island.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 107057)

## AUTHORS

Direct Submission

## JOURNAL

Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

## COMMENT

requests: clonerequest@sanger.ac.uk  
On Mar 19, 2000 this sequence version replaced gi:6523711.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
EM: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e. phred quality >= 30);  
an attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP5-107712 is from  
the library RPCI-5 constructed by the group of Pieter de Jong. For  
further details see  
http://www.chori.org/bacpac/home.htm

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone  
RP5-107712 It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.

The true right end of clone RP5-107712 is at 107057 in this  
sequence. The true right end of clone RP4-718P11 is at 100 in this  
sequence.

## FEATURES

Location/Qualifiers

1..107057  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone="RP5-107712"  
/clone\_lib="RPCI-5"  
repeat\_region 286..490

```
repeat_region /note="LIP44 repeat: matches 5942. .6146 of consensus"
misc_feature 1467. .1494
/note="14 copies 2 mer ac 89% conserved"
repeat_region 1889. 2354
/note="match: STS: Em: AAL59919"
misc_feature 2112. 2748
/note="match: STS: Em: AA005104"
repeat_region 2399. 2464
/note="MER5A repeat: matches 41. .105 of consensus"
misc_feature 2513. 2619
/note="match: STS: Em: HSPF04E9 Em: Z94369"
/note="complement(2851. .3124)"
misc_feature /note="match: STS: Em: G29540 Em: T98667"
/note="complement(2856. .3124)"
repeat_region /note="match: STS: Em: G25883"
3615. 3653
/note="3 copies 13 mer 94% conserved"
repeat_region 4108. 4247
/note="MER5A repeat: matches 44. .184 of consensus"
/note="complement(4488. .5147)"
misc_feature /note="match: GSS: Em: AQ262066"
4541. 4855
repeat_region /note="AluSc repeat: matches 1. .309 of consensus"
complement(4979. .5508)
misc_feature /note="match: GSS: Em: AQ664680"
complement(4996. .5415)
misc_feature /note="match: STS: Em: N30188"
5315. 5364
/note="25 copies 2 mer tg 84% conserved"
repeat_region 5318. 5365
/note="12 copies 4 mer tgt 85% conserved"
repeat_region 5489. 5566
/note="LTR16A repeat: matches 378. .450 of consensus"
repeat_region 5567. 5760
/note="MER63A repeat: matches 1. .209 of consensus"
repeat_region 5761. 6016
/note="LTR16A repeat: matches 95. .378 of consensus"
6044. 6288
/note="L1MC5 repeat: matches 7695. .7937 of consensus"
misc_feature 7638. 8163
/note="match: GSS: Em: B52469"
misc_feature 7685. 8134
/note="match: GSS: Em: AQ667183"
repeat_region 8270. 8512
/note="HAL1 repeat: matches 162. .421 of consensus"
repeat_region 8329. 8462
/note="67 copies 2 mer aa 59% conserved"
repeat_region 8970. 9273
/note="HAL1 repeat: matches 680. .997 of consensus"
10789. 11086
/note="AluJb repeat: matches 1. .297 of consensus"
repeat_region 11170. 11523
/note="THL1B repeat: matches 1. .364 of consensus"
repeat_region 11578. 11752
/note="MER86 repeat: matches 2. .195 of consensus"
complement(11947. .12544)
misc_feature /note="match: STS: Em: AL110056 Em: HS75707T"
12307. 12526
/note="L1MEC repeat: matches 1192. .1424 of consensus"
repeat_region 12529. 13073
/note="L1MA5A repeat: matches 3652. .4242 of consensus"
repeat_region 13076. 13330
/note="L1MA5A repeat: matches 4230. .4487 of consensus"
repeat_region 13331. 13638
/note="AluSx repeat: matches 1. .307 of consensus"
repeat_region 13639. 15444
/note="L1MA5A repeat: matches 4487. .6295 of consensus"
repeat_region 15775. 15814
/note="10 copies 4 mer tgg 95% conserved"
misc_feature /note="complement(16370. .16537)"
complement(16370. .16537)
repeat_region 17517. 17817
/note="LTR16C repeat: matches 82. .385 of consensus"
```

alignment\_scores:  
Quality: 321.00  
Ratio: 5.350

Length: 60  
Gaps: 0

4'

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-863-824-2 x HSJ107712 ..

Align seg 1/1 to: HSJ107712 from: 1 to: 107057

```
47 AsnAsnLeuAsnValGlySerAspThrThrSerGluThrSerPheSerLe 63
|||||
106877 AATAACCTCAACGTGGGAAGTGACACCAATCAGAAACACGCTTTCTCT 106926
|||||
63 uSerLysGluAlaProArgGluHisLeuAspHisGlnAlaHisGlnP 80
|||||
106927 CTCRAAAGAACCAAGGAGCATCTGGACCACCAAGCTGCACACCAAC 106976
|||||
80 roPheProArgProArgPheArgGlnGluThrGlyHisProSerLeuGln 96
|||||
106977 CCTTCCCAAGACGGGATTCGACAGAAGACGGGGACCCCTTCATGCAA 107026
|||||
97 ArgAspPheProArgSerPheLeuLeuAsp 106
|||||
107027 AGAGATTCCCAAGATCCCTTCTCTTGAT 107056
|||||
```

seq\_name: gb\_sts:HS107712S

seq\_documentation\_block: 523 bp DNA linear STS 24-AUG-1999  
LOCUS HS107712S  
DEFINITION H.sapiens STS from genomic clone 107712, sequence tagged site.  
ACCESSION AL110084  
VERSION AL110084.1 GI:5777434  
KEYWORDS STS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 523)  
AUTHORS Deloukas,P., O'Neill,L., Holden,J., Mistry,D., Huckle,E., Taylor,R.  
and Hunt,S.

Direct Submission

Submitted (24-AUG-1999) E-mail contact: humquery@sanger.ac.uk  
Marker: STJ107712SP6 (Primer A : GGAAATCTCTTCAATGAGG; Primer B :  
ACCTCAACGGGGGAAGGAC; amplicmer size : 156 bp) is from sequence  
generated from the SP6 end of PAC 107712. 107712 is part of the  
bacterial clone contigs constructed by the Chromosome 20 Mapping  
Library.(http://www.sanger.ac.uk/HGP/Chr20/) 107712 is from the  
group constructed at the Roswell Park Cancer Institute by the  
group of Pieter de Jong.  
For further details see http://bacpac.med.buffalo.edu/.

Location/Qualifiers  
1..523  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone="107712"  
BASE COUNT 150 a 102 c 116 g 155 t  
ORIGIN

FEATURES

source

alignment\_scores:  
Quality: 292.00 Length: 54  
Ratio: 5.407 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-863-824-2 x HS107712S/rev ..

Align seg 1/1 to reverse of: HS107712S from: 1 to: 523

```
47 AsnAsnLeuAsnValGlySerAspThrThrSerGluThrSerPheSerLe 63
|||||
162 AATAACCTCAACGTGGGAAGTGACACCAATCAGAAACACGCTTTCTCT 113
|||||
63 uSerLysGluAlaProArgGluHisLeuAspHisGlnAlaHisGlnP 80
|||||
```

```
|||||
112 CTCCAAGAGACCAAGGAGCATCTGGACCACCAAGCGCTGCACCAAC 63
|||||
80 roPheProArgProArgPheArgGlnGluThrGlyHisProSerLeuGln 96
|||||
62 CCTTCCCAAGACCGGATTCGACAGAAGACGGGGACCCCTTCATGCAA 13
|||||
```

seq\_name: gb\_ro:MUSTSPIA

seq\_documentation\_block: 4339 bp mRNA linear ROD 09-NOV-1994  
LOCUS MUSTSPIA  
DEFINITION Mouse thrombospondin 1 mRNA, complete cds.  
ACCESSION M87276  
VERSION M87276.1 GI:202196  
KEYWORDS thrombospondin 1.  
SOURCE Mus musculus (strain CD-1) embryo kidney cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 4339)  
AUTHORS Laherty,C.D., O'Rourke,K., Wolf,F.W., Katz,R., Seldin,M.F. and  
Dixit,V.M.

Characterization of mouse thrombospondin 2 sequence and expression  
during cell growth and development

J. Biol. Chem. 267 (5), 3274-3281 (1992)

92147683 Location/Qualifiers

1..4339  
/organism="Mus musculus"  
/strain="CD-1"  
/db\_xref="taxon:10090"  
/tissue\_type="kidney"  
/dev\_stage="embryo"  
212..3724  
/codon\_start=1  
/product="thrombospondin 1"  
/protein\_id="AA53063.1"  
/db\_xref="GI:567240"  
/translation="MELLRLGLVFLHLMCGSNRIPESGGDNGVDFIFELIGARRGP  
GRPLVKGDLSSPAFENANLIPAPDDKFODLLDAVMDKGFIFLASLRQMKTRG  
TLIAVERKNTQIFSVNSGKAGTLDLSLPGKQVSVVEALLATQOMKSLTFLV  
QEDRAOLYIDCKMESAELDFIQSIFTRDLASVARLRVAKGDVNDNFQGVQLNRFV  
FGTTPEDILRNKCGSSSTNVLITLDNNVNGSSPAIRTNVIGHKTKDLQAIGLSCDE  
LSSMVLKGLRTIVTTLQDSIRKVTENRELVSELKRPPLCFHNGVQTKNNEWTVD  
SCTECHQNSVTICKKVCPIMPCSNATVPDGECCPCRPWSDSADGGSPWSEWTSVS  
ATCGNGIQGRSCDSLNRCESSVQTRTCHTQCDKRFKQDGGWSPWSDSVTC  
CDGVITRILCNPSQPMNGKPCGEARETRACKKDCPIIDGCLSNPCFAGAKTSTP  
GGVQRRSLCNPTPQFGGKDCQDVTEQVNCNKDCPIIDGCLSNPCFAGAKTSTP  
DGSWKCGACPPGYSNGIQCQDVDECKEVPDCAFHNHGRCKNTDPGYNCLPCPPRF  
TGSOPEGRGVEHAMANKQVCPNRPCTDGTDCNKNKAKNLYLGHVSDPMYRCEKPGY  
AGNGIICGEDTDLDGWPNENLVANATYHCKKDCNCPNHNPDQADTKNGEGDACA  
DDNDKIIPDDRNCDFHYNPAQYDTRDDVDGDCNCPNHNPDQADTKNGEGDACA  
VDIDGGLINERDNCQYVNVQDRTDMDGVGDCQDCNCPNHNPDQADTKNGEGDACA  
DNNQDIDEGHQNHLNCPVNPNAOAHDKGKGDACDHDNDGIPDDRNCRLVP  
NPDKSDSDGGRGACKDFDHDNVDPIDDIICPENFDISETDPRRFQMLPDKGTSQ  
NDPNVVRHOGKELVQTCPLGLVDFEYDFNAVDFSGTFFINTERDDYAGLVFGYQ  
SSRFYVVMKQYTSYDTPNTPRAQGYSLGSVKVYVNSTTGPGEHLRNALHWTGTPG  
QVRLWHDPRHIGWKDFYRWRLSHRPKTGYIRVVMYEGKKIMADSGPIYDKTYAGG  
RLGLVFVSQBMVFFSDMKYBCRDS"

BASE COUNT 1109 a 1201 c 1144 g 885 t  
ORIGIN

alignment\_scores:

Quality: 145.50 Length: 328

Ratio: 0.951 Gaps: 17

Percent Similarity: 46.646 Percent Identity: 24.085

alignment\_block:

```

US-09-863-824-2 x MUSTSP1A ..
Align seg 1/1 to: MUSTSP1A from: 1 to: 4339

15 LeuLeuLeuThrLeuLeuHisIleThrValLeuArgGlySerGlyAlaAla.A 31
   :::::::::::::::::::::
923 GTCCCTCTTACCTTGACAACAACGCTGTGAACGGTTCAGCCCTGCTAT 972

31 spGlyProAspAlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsn 47
   :::::::::::::::::::::
973 CCGCACCACTACATCGGCC.....ACAAACAAGAGGACCTCC 1010

48 AsnLeuAsnValGlySerAspThrThrSerGluThrSerPheSerLeu.S 64
   :::::::::::::::::::::
1011 AAGCTATCTGTGGCTCTCCTGTGATGAATATCCAGCATGCTCTGGAA 1060

64 erLysGluAlaProArgGluHisLeuAspHisGlnAlaAlaHisGlnPro 80
   :::::::::::::::::::::
1061 CT...GAAGGGCTCGGCACCATCTGTACCACCTCTGCAGGACAGCATCCG 1107

81 PheProArgProArgPheArgGln.....GluThrGlyHi 92
   :::::::::::::::::::::

1108 AAAAGTGACGAAGAACAAGACAGAGCTGCTGAGTGAGTGAAGCGGCTC 1157

92 sProSerLeuGlnArg.....AspP 99
   :::::::::::::::::::::

1158 CCTCTGCTTTCACAATGGAGTCCAGTACAAGACAAGAGGAGTGACT 1207

99 heProArgSerPheLeuLeuAspLeuPro.AsnPheProAspLeuSerLy 115
   :::::::::::::::::::::
1208 GTAGACAGTTCACACAGAGTGTCACTGCCAGAACTCGTTTACCATCTGCAA 1257

115 sAlaAspIleAsnGlyGln**ProAsnIleGlnValThrIleGluValV 132
   :::::::::::::::::::::
1258 AAAGTGTCTGTCCCATCATCGCTGCTCCACGCCACAGTT..... 1300

132 alAspGlyProAspSerGluAlaAspLysAspGlnHisProGluAsnLys 148
   :::::::::::::::::::::
1301 .....CCTGATGCTGAATGC.....TGCCACGGTGCTGG 1330

149 ProSerTrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSe 165
   :::::::::::::::::::::
1331 CCCAGCGACTGTCT...GACGATGGCTGGTCTCCCTGCTGAGTGGAC 1377

165 rLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAsp.....Tyr***T 180
   :::::::::::::::::::::
1378 CTCCTGCTCTGCCACATGTGCAATGGAAATTCAGCAACGTGCTGCTTCT 1427

180 yrAspSerThrSerAsp..... 185
   :::::::::::::::::::::
1428 GTGACAGCCTCAACAACAGATCGGAGGCTCTTCGGTACAGCAGGACC 1477

186 .....AspSerAsnPheLeuAsnProProArgGlyTr 196
   :::::::::::::::::::::
1478 TCCACATTCAGGAGTGTCAAAAAGATT...AAACAGGATGCTGGCTG 1524

196 pAspHisThrAlaProGlyHisArgThrPheGluThr..... 208
   :::::::::::::::::::::
1525 GAGTCACCTGGTCTCCATGCTGCTCTGTTCTGTGACCTGTGGTGACGTG 1574

209 .....LysAspGlnProGluTrpAspSer 216
   :::::::::::::::::::::
1575 TGATCACAAGGATCGTCTCTGCAACTCCCCCAGCCGCCAGATGAACGGG 1624

217 ThrAspGlyGluGlyAsp..... 222
   :::::::::::::::::::::
1625 AAGCCCTGTGAAGTGAAGCCCGGAGACCAAGCCTGCAAGAAAGACGC 1674

223 .....TrpSerLeuTrpSerValCysS 230
   :::::::::::::::::::::
1675 CTGCCCAATTATGAGGCTGGGTCTCCTGTGTCACCATGGGACATCTGCT 1724

230 erValThrCysGlyAsnGlyAsnGlnLysArgThrArgSerCys..... 244

```

```

1725 CTGTCACTGTGGAGGAGTGCGACAGACCGACCGACTCTGTAAACAC 1774
   :::::::::::::::::::::
245 .....GlyTyrAlaCysThr.....AlaThrGluSe 253
   :::::::::::::::::::::
1775 CCCACACCCAGTGTGGAGGCAAGACTGTGTGGCGATGTGACAGAAAA 1824

253 rArgThrCysAspArgProAsnCysPro 262
   :::::::::::::::::::::
1825 TCAAGTTTGCACCAAGCAGGACTGCCCA 1852

seq_name: gb_htg:AL626774

seq_documentation_block:
LOCUS AL626774 199982 bp DNA linear HTG 21-OCT-2001
DEFINITION Mus musculus chromosome 4 clone RP23-198015, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL626774
VERSION AL626774.8 GI:16304784
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 199982)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 21, 2001 this sequence version replaced gi:16215137.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BM198015
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194180 bases at least Q40
Consensus quality: 195534 bases at least Q30
Consensus quality: 196426 bases at least Q20
Insert size: 197682; sum-of-contigs
Quality coverage: 8.11x in Q20 bases; sum-of-contigs Quality
coverage: 8.43x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
    Location/Qualifiers
        1..199982
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="4"
            /clone_lib="RPCI-23"
            /clone="RP23-198015"
            1..8011
                /note="assembly_fragment:00065"
                fragment_chain:1
                    clone_end:SP6
                    vector_side:left
                        8112..15629
                            /note="assembly_fragment:01193"
                            fragment_chain:1
                                15730..19452
                                    /note="assembly_fragment:04957"
                                    fragment_chain:1
                                        19553..21702

```

```
misc_feature /note="assembly_fragment:00304
fragment_chain:2"
21803..31551
/note="assembly_fragment:03726
fragment_chain:2"
31652..38994
/note="assembly_fragment:05064
fragment_chain:2"
39095..52821
/note="assembly_fragment:03824
fragment_chain:2"
52922..61341
/note="assembly_fragment:01209
fragment_chain:2"
61442..65388
/note="assembly_fragment:01814
fragment_chain:2"
65489..69123
/note="assembly_fragment:00089
fragment_chain:2"
69224..71773
/note="assembly_fragment:04270
fragment_chain:2"
71874..79236
/note="assembly_fragment:01456
fragment_chain:2"
79337..81353
/note="assembly_fragment:03559
fragment_chain:2"
81454..83779
/note="assembly_fragment:04965
fragment_chain:3"
83880..94295
/note="assembly_fragment:03552
fragment_chain:3"
94396..112245
/note="assembly_fragment:00547
fragment_chain:3"
112346..118182
/note="assembly_fragment:03745
fragment_chain:3"
118283..126184
/note="assembly_fragment:04640
fragment_chain:4"
126285..159918
/note="assembly_fragment:00131
fragment_chain:4"
160019..169858
/note="assembly_fragment:04645
fragment_chain:5"
169959..184421
/note="assembly_fragment:01090
fragment_chain:5"
184522..191988
/note="assembly_fragment:04278
fragment_chain:6"
192089..196969
/note="assembly_fragment:01779
fragment_chain:6"
197070..199982
/note="assembly_fragment:03639
fragment_chain:6
clone_end:17
vector_side:right"
BASE COUNT 50475 a 50251 c 48629 g 48320 t 2307 others
ORIGIN
```

```
alignment_scores:
  Quality: 144.50      Length: 236
  Ratio: 1.290        Gaps: 11
Percent Similarity: 47.458 Percent Identity: 27.966
```

```
alignment_block:
US-09-863-824-2 x AL626774/rev ..
Align seg 1/1 to reverse of: AL626774 from: 1 to: 199982
73 AsphHisGlnAlaAlaHisGlnPropheProArgProArgPheArgGlnG1 89
||||:||||| ||| |||
60619 GACCAACAGGCTGTGCATCTCCTGTGGAACCTCCGCCCGCGCA 60570
89 uThr.....GlyHisProSerLeuGlnA 97
||||:||||| ||| |||
60569 CACACACTAGCCACCACAGTTCTGCTCAATGGCCACCCCGCTTG.... 60524
97 rgAspPheProArgSerPheLeuLeuAspLeuProAsnPheProAspLeu 113
||||:||||| ||| |||
60523 .....TGTCTCCCATGCAG 60509
114 SerLysAlaAspIleAsnGlyGln**ProAsnIleGlnValThrIleG1 130
||||:||||| ||| |||
60508 TGAAGGCCAGTGTGCTAGAAATGGGGTCCCTGGGCCCATGCTCATCATCT 60459
130 uValValAspGlyProAspSerGluAla...AspLysAspGlnHisProG 146
||||:||||| ||| |||
60458 TGTGCCAATGGACCCAGCAGCGCGGAATGCAGTGTGGCGGGTCC 60409
146 luAsnLysProSerTrpSerValProSerProAspTrpArgAlaTrpTrp 162
||||:||||| ||| |||
60408 AGCGTGGGCCACGTCGCGCAGGTGCCTCA.....CGGATACCCGTG 60368
163 GlnArgSerLeuLeuAlaArgAlaAsnSer.GlyAspGlnAspTyr* 179
||||:||||| ||| |||
60367 AGTCAGCAATCTCGATGTCGCCGGGTGAGTGTGGTGTGAGG..... 60325
60324 .....GTAGCAGGGGTAGCCCTGGCTTGGAGACTTGCATGGC 60286
196 TrpAspHisThrAlaProGlyHisArgThrPheGluThrLysAspGlnPr 212
||||:||||| ||| |||
60285 CTCTCACACTTC.....CC 60272
212 oGluTyrAspSerThrAspGlyGlu...GlyAspTrpSerLeuTrpServ 228
||||:||||| ||| |||
60271 TTCCTCATAGCCACTGACGCAAGTGGGGCGCGTGGAAACCGGTGAGGCC 60222
228 alCysSerValThrCysGlyAsnGlyAsnGlnLysArgThrArgSerCys 244
||||:||||| ||| |||
60221 TGTGCTCAAGACGTGTGACAGGGCTGGCAACGCCGCTCCGCATGTGC 60172
245 .....GlyTyrAlaCys.....ThrAlaThrGluSe 253
||||:||||| ||| |||
60171 CAGGCTTCTGGCACACAGGGCTTACCTTGGCAGGGCAGAGGAGAGGT 60122
253 rArgThrCysAspArgProAsnCysProGlyIleGluAspThrPheArgT 270
||||:||||| ||| |||
60121 GAAACCTGTCAGTGAGAGAGGTGTCAGGTATT..... 60088
270 hrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAla 286
||||:||||| ||| |||
60087 ..GCTTATACAAACCTGGGGCTGGGGGCTGGGGGACAGATATAGCCCA 60040
287 ThrLys 288
||||:||||| ||| |||
60039 GCTAAA 60034
seq_name: gb_pr:AB005298
seq_documentation_block:
LOCUS AB005298
DEFINITION Homo sapiens BAI 2 mRNA, complete cds.
ACCESSION AB005298
VERSION AB005298.1 GI:3021698
KEYWORDS BAI 2.
```



SOURCE Homo sapiens brain cDNA to mRNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 5399)  
 AUTHORS Nakamura, Y.  
 JOURNAL Direct Submission  
 Submitted (28-JUN-1997) Yusuke Nakamura, The Inst. of Medical  
 Science, The University of Tokyo, Lab. of Molecular Medicine, Human  
 Genome Center, 4-6-1, Shirokanedai Minato-ku, Tokyo, Minato-ku,  
 Tokyo 108, Japan (E-mail: yusuke@ims.u-tokyo.ac.jp,  
 Tel: 81-3-5449-5372, Fax: 81-3-5449-5433)  
 REFERENCE 2 (sites)  
 Shirauchi, T., Nishimori, H., Ichise, H., Nakamura, Y. and Tokino, T.  
 Cloning and characterization of BAI2 and BAI3, novel genes  
 homologous to brain-specific angiotensin inhibitor 1 (BAI1)  
 Cytogenet. Cell Genet. 79 (1-2), 103-108 (1997)

FEATURES

Location/Qualifiers

1..5399  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="lp35"  
 /tissue\_type="brain"  
 379..5097  
 /gene="BAI 2"  
 379..5097  
 /gene="BAI 2"  
 /codon\_start=1  
 /protein\_id="BAA25362.1"  
 /db\_xref="gi:3021699"  
 /translation="MTAPCLLSVLSRLATAPDPAPSAACSLASGLVYGAFLSDP  
 LFPTIASCSWLENPDPTKSYLRFNRQBOVCAHFAFRLPLDHYLVNFTCLRPSP  
 EEAQAQSEVREEREEAEEAAGLELCSGSPTEFLHDFNFVQLCLASBPSEAPRL  
 LAPALAFREVEVLINNNSSQFTGVCWRSECEGRAGACGAFPGSCGPEAG  
 AGSTTTSPGPAATLSNALVPGGAPPADLHSGSNDLFTTEMRYGEEPEEPK  
 VTKQWPSADPGLYMAQTDGPAEAEWSPVSLTCGQGVTRVRSVYSGPYGLCS  
 GLRTRPCNNASATCPVHGVWEEMGWSLCSRSGRSRMRCTVPPHGGKACEGP  
 ELQTKLCSMAACPVEGWLEWPGWPCSCANGTQORSRKCSVAGPAWACTGALT  
 TRECNECPATDSKWPNAWSLCSKTDGTGWQRRFRMCAATGTGQYPCGTEGEVK  
 PCSERCAFHEMCDEYVLMWTKAAAGELIYNKPPNAGSASRSLCLASQAQVAY  
 WGLPFSFARISHEYRILYLREHLAKQRLAGEMSQVRSQELRLARTYSVGD  
 LFSVDILRNVDYTKRATYVPSADVDQRFVQFVPMVDAENKEKWDAAQQYSPGSVHL  
 RVVEDFLHLVGDALKAFQSLIVTDNLVISTOREPVSASVSDITFPMRGRGMKDW  
 RHSEDFLPLKXEVLSLSPGKPATSGAAGPCRGCTGCTVPGCHSHQRLPADPDE  
 SSYFVIGAVLRTGLILPPRPPLAVTSRVTVTPPTQPPAEPLITVELSYIING  
 TDPHCSWDYSRADASSGDWDTENCQTLTOAAHRCQCQHLSTFAVLQPKDLTL  
 ELAGSPSVPLVIGCAVSCMLTLTLLAYAFWREFTKRSRIILNFCILSLASNLIL  
 VQSRVLSKGVCTMTAFLHFFLSFCWLTAEQSYLAVIGRMRTLVRKRFCLIG  
 WGLPALVVASVGTGRTKGYGTSVYCWLSLEGLLYAFVGAATVLYNMLIGLIIFN  
 KLMDRGISDSKQKQAGSERCPWASLLILPCSGAGVPSPLSSASARNASLWSSC  
 VYLLPLATWMSAVLAMDTRRSVLVQALFAVNSAQGFVITAVHCFURREVODVVKCQ  
 MGVCRLADSSPSCNGKQQLIDFEDKVDLACQVLFKEVNTCPNSTITGLSLRL  
 SLDEDEEPKSLGVLPGLSPGLNLPVMAASPGLEPPPOEANPVMCMGEGGL  
 RQDLDTLWLRPEPGSEGDMVLPRLTSLQPGGGGGEDAPRARPECTPRAAKTVA  
 HTEGYPSPLSVDHSLGLGAYGSLQNPYGMTFPPPTPSAROVPEGERSRTPRT  
 VPGSTMKGLSLERKRLRYSDLDFFVMTFRKHSLEYHLNOKFHTFDYRSQSTAKRE  
 KRWVSYSGGAERSVCTDKPSPGERPSLSQHRHRSQSWSTFKSMTLGSLPPKRLTL  
 HRAAAWEPTPPDGDQFEV"

gene

CDS

polyA\_site

BASE COUNT

ORIGIN

alignment\_scores

Quality:

Ratio:

Percent Similarity:

alignment\_block:

US-09-863-824-2 x AB005298

Align seg 1/1 to: AB005298 from: 1 to: 5399

114 SerLysAlaAspIleAsnGlyGln\*\*ProAsnIleGlnValThrIleG1 130  
 1120 GCTGAGGCCCATTTGCACCTCGGGAGCAGCAATGATCTGTTCAACACGA 1169  
 130 uValValAspGlyProAspSerGluAlaAspLysAspGlnHisProGluA 147  
 1170 GATGAGATATGGTGAGGACCGGAA.....GAGGAACCGAAAG 1207  
 147 snLysProSerTrp.....SerValProSerPro..... 156  
 1208 TGAACACCCAGTGGCCGAGGTCTGCAGATGAGCCTGGGCTATACATGGCG 1257  
 157 .....AspTrpArgAlaTrpTrpGlnArgSe 165  
 1258 CAGACAGCGACCGCGGCTGAGGAGTGTCCCGCTGGAGCGGTGTGTTTC 1307  
 165 rLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAspTyr\*\*TyrAsps 182  
 1308 CTTGACG.....TGTGGCAGGTCTCCAGGTGGGACCGCTCCT 1348  
 182 erThrSerAspAspSerAsnPheLeu.....Asn 191  
 1349 GTGTGTCTCTCCCTATGGGACCTGTGTGACGGGGCCCTCGGGGAGACC 1398  
 192 ProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGlu... 207  
 1399 AGGCCCTGCAACAATTCAGCCACCTGCCACGTGCACGCGGTGTGGAGGA 1448  
 207 ..... 207  
 1449 GTGGGGTCTCTGGAGCCTGTGCTCCCGCAGCTGCGGGGGGGTCCCGGA 1498  
 208 .....ThrLysAspGlnProGlyTyrAspSerThrAspGlyGlu 220  
 1499 GCGCGATGCGGACCTGTGCTGCCCGCCAGCAGCGCGGACGCGCTGCGAG 1548  
 221 Gly..... 221  
 1549 GGTCTGAGCTGACAGACTAGCTGTGCAATGAGTGTGCTGCCCGGCTGGA 1598  
 222 .....AspTrpSerLeuTrpSerValCysSerValThrCysG 234  
 1599 AGGCCAGTGTAGAAATGGGGTCTCTGGGGCCCATGCTCCACGCTCTGTG 1648  
 234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAla..... 247  
 1649 CCAATGGGCCCAACACAGCGAGCGGAAGTGCAGCGTGGCGGGCCAGCC 1698  
 248 .....CysThrAla.....ThrGluSerArgThrCysAspArgPr 259  
 1699 TGGGCCACATGACGGGTGCGCTCCTGACACCGCGGAGTGCAGCAACCT 1748  
 259 oAsnCysProGlyIleGluAspThrPheArgThrAlaAlaThrGluValS 276  
 1749 CGAGTGGCGGCCACTGATGCAAGTGG..... 1776  
 276 erLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheGlu... 291  
 1777 .....GGGCCATGGAATGCGTGGAGCCTGTGCTCTAAG 1809  
 292 ...ValAspThrAspSerCysGluArgTrpMetSerCysLysSerGluPh 307  
 1810 ACGTGTACACAGCGCTGCAGCGCTTCGCGCATGTGCGAGGACGCGG 1859  
 307 eLeuLysLysTyrMethIleLysValMetAsnAspLeuProSerCysProC 324  
 1860 CAGCGAGGCTAC..... 1876  
 324 ySerTyrProThrGluValAlaTyrSerThrAlaAsp..... 336

seq_name:	gb_pat:	E21623	seq_documentation_block:	5412 bp	DNA	linear	PAT 07-FEB-2001
1877	GCAGAGCCACCGAGGAGGAGGTGAACCCCTGTGTAGTGAGAGAGGTGTCCA	1926					
337	ILlPheAspArgTleLysArgLysAspPhe.....ArgTrpLys	349					
1927	GCCTTCATGAGATGTCAGGGATGAGTACGTCATGCTGATGACGTGGAA	1976					
349	sAspAlaSerGlyProLysGluLysLeuGlyLysLeuLleTyrLysProThrAla.	365					
1977	GAAGGCAGTCGCTGGCGAGATCATCATACAAAGTGCCTCCCGCAATGCCT	2026					
366	.....ArgTyrCysIleArgSer.....	371					
2027	CAGGGTCTGCCACCGCGCTGTCTCTCAGTCCCAAGCGTGGCGTAC	2076					
371	.....	371					
2077	TGGGGGTGCCAGCTTTGCTGCTGCATCTCCATGAGTACGCTACCT	2126					
372	MetLeuSerLeuGluSerThrThrLeuAlaAlaGlnHisCysCysTyrG	388					
2127	GTATCTGTCTACTTAGGAGGACCACTGGCCAGGGCAGCGCATGCTGGCAG	2176					
388	lyAspAsnMetGlnLeuIleThrArgGlyLysGlyAlaGlyThrProAsn	404					
2177	GCAGGGCAGTCGTCAGGTGGTGGC.....AGCCTGCAGGAG	2214					
405	LeuIleSer.....ThrGluPheSerAlaGluLeuHisTyrLysValas	419					
2215	CTACTGCCCGGCACCTACTAGTAGGGGACCTCTCTCTCTGTGGGA	2264					
419	pValLeuProTrpIleIleLysGlyAspTrpSerArgTyrAsnGluA	436					
2265	CATTCTG.....A	2272					
436	laArgProProAsnAsnGlyGlnLysCysThrGluSerProSerAspGlu	452					
2273	GGATGTCACTGACACCTTTAAGAGGGCCACCTACGTGCCCTCGGCTGAT	2322					
453	AspTyrIleLysGlnPheGln	459					
2323	GATGTGAGCGCTCTTCCAG	2343					
seq_name:	gb_pat:	E21623	seq_documentation_block:	5412 bp	DNA	linear	PAT 07-FEB-2001
LOCUS	E21623						
DEFINITION	Human BAI gene and utilization thereof.						
ACCESSION	E21623						
VERSION	JP 1999032766-A/2.						
KEYWORDS	unidentified.						
SOURCE	unidentified.						
ORGANISM	unclassified.						
REFERENCE	1 (bases 1 to 5412)						
AUTHORS	Ryuji, T.Y.N.N.						
TITLE	Human BAI gene and utilization thereof						
JOURNAL	Patent: JP 1999032766-A 2 09-FEB-1999;						
COMMENT	OTSUKA PHARMACEUT CO LTD						
	OS Unidentified						
	PN JP 1999032766-A/2						
	PD 09-FEB-1999						
	PF 16-JUN-1997 JP 1997176485						
	PR						
	PI RYUJI TOKINO, YUSUKE NAKAMURA						
	PC C12N15/09, A61K38/00, C07H21/04, C07K14/47, C12N15/00, A61K37/02						
	CC Strandedness: Single;						
	CC Topology: Linear;						
	FH Key						
	FT CDS						
FEATURES	Location/Qualifiers						
source	1..5412						
	/organism="unidentified"						
	/db_xref="taxon:32644"						

```
292 ...ValAspThrAspSerCysGluArgTyrMetSerCysLysSerGluPh 307
||||| :|||: :|||: :|||: :|||:
1810 ACGTGTGACACAGCGTGGCAGCGCGTTCGGCATGTGCCAGGCCAGGG 1859
||| :|||: :|||: :|||: :|||:
307 eLeuLysLysTyrMethHisLysValMetAsnAspLeuProSerCysProC 324
::: ||| :|||: :|||: :|||: :|||:
1860 CAGCGAGGCTAC.....CCCT 1876

324 ySerTyrProThrGluValAlaTyrSerThrAlaAsp..... 336
||||| :|||: :|||: :|||: :|||:
1877 GCGAGGCGACCGAGGAGGTGAAGCCTTGTAGTGAGAGAGGTGCCA 1926

337 IlePheAspArgIleLysArgLysAspPhe.....ArgTyrPly 349
||| :|||: :|||: :|||: :|||:
1927 GCCTTCATGATGTCAGGGATGATGATGATGATGATGATGATGATGAA 1976

349 sAspAlaSerGlyProLysLysLysLysLysLysLysLysLysLysLys 365
| :|||: :|||: :|||: :|||: :|||:
1977 GAAGGAGCTGCTGGCGAGATCATCTACAACAAGTGCCCGCGAATGCCT 2026

366 .....ArgTyrCysIleArgSer..... 371
||| :|||: :|||: :|||: :|||:
2027 CAGGGTGTCCAGCGCGCGTGTCTCTCTCAGTGCCTCAAGCGGTGGCTAC 2076

371 ..... 371

2077 TGGGGGCTGCCAGCTTGTGCTGCTGCATCTCCATGATACCGCTACCT 2126

372 .MetLeuSerLeuGluSerThrThrLeuAlaAlaGlnHisCysCysTyrG 388
||||| :|||: :|||: :|||: :|||:
2127 GTATCTGTCTACTTAGGAGCACCTTGGCCAAAGGGCAGCGCATGCTGCG 2176

388 lyAspAsnMetGlnLeuIleThrArgGlyLysGlyAlaGlyThrProAsn 404
||||| :|||: :|||: :|||: :|||:
2177 GCGAGGCGATGTGCGAGGTGTGCGC.....AGCCTGCGAGGAG 2214

405 LeuIleSer.....ThrGluPheSerAlaGluLeuHisTyrLysValas 419
||||| :|||: :|||: :|||: :|||:
2215 CTACTGCGCGCGGCACCTACTATAGTGGGACCTGCTCTTCTGTGTGA 2264

419 pValLeuProTyrPheIleCysLysGlyAspTyrSerArgTyrAsnGluA 436
||||| :|||: :|||: :|||: :|||:
2265 CATCTGT.....A 2272

436 laArgProProAsnAsnGlyGlnLysCysThrGluSerProSerAspGlu 452
::: :|||: :|||: :|||: :|||:
2273 GGAATGTCTACTGACACCTTTAAGAGGGCCACCTACGTGCCCTCGGCTGAT 2322

453 AspTyrIleLysGlnPheGln 459
||| :|||: :|||: :|||: :|||:
2323 GATGTGCGAGCGCTCTTCCAG 2343
```

seq\_name: gb\_pat:E21622

```
seq_documentation_block:
LOCUS      E21622              5535 bp      DNA      linear      PAT 07-FEB-2001
DEFINITION Human BAI gene and utilization thereof.
```

ACCESSION E21622

VERSION E21622.1 GI:13023632

KEYWORDS JP 199032766-A/1.

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 5535)

AUTHORS Ryuji.T.Y.N.N.

TITLE Human BAI gene and utilization thereof

JOURNAL Patent: JP 199032766-A 1 09-FEB-1999;

OTSUKA PHARMACEUT CO LTD

COMMENT OS Unidentified

PN JP 1999032766-A/1

PD 09-FEB-1999

PF 16-JUN-1997 JP 1997176485

```
PR      RYUJI TOKINO,YUSUKE NAKAMURA
PI      C12N15/09,A61K38/00,C07H21/04,C07K14/47,C12N15/00,A61K37/02 CC
PC      Strandedness: Single;
CC      Topology: Linear;
FH      Key      Location/Qualifiers
FT      CDS      184..4938.

FEATURES             Location/Qualifiers
     source           1..5535
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      944 a 1937 c 1780 g 874 t
ORIGIN
```

```
alignment_scores:
Quality: 142.00      Length: 542
Ratio: 0.664        Gaps: 26
Percent Similarity: 39.483 Percent Identity: 22.509

alignment_block:
US-09-863-824-2 x E21622 ..
Align seg 1/1 to: E21622 from: 1 to: 5535
```

```
2 ValArgLeuAlaAlaGluLeuLeuLeuGlyLeuLeuLeuLeuTh 18
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
214 GTCGTGGATCTCGCCCGCTGCTACTGCTG.....CTGCTGTGCTGGG 257

18 rLeuHisIleThrValLeuArgGlySerGlyAlaAlaAspGlyProAspA 35
::: :||| :||| :||| :||| :||| :||| :||| :||| :|||
258 ACGCGCGCGCGCGCGCGCGCGCGAGCAGCGCGCGCGCGCGCGCGGAGC 307

35 laAlaAla.....GlyAsnAlaSerGlnAla 43
::: ||| :||| :||| :||| :||| :||| :||| :||| :|||
308 CGTGCAGCACGCTGGTGCAGGAAAGTTCTTCGGCTACTTCTCGCGCGCC 357

44 GlnLeuGlnAsnAsnLeuAsnValGlySerAspThrThrSerGluThrSe 60
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
358 GCC.....GTGTTCCGCGCAACGCGCTCGCGCTGCTC 389

60 rPheSerLeuSerLysGluAlaProArgGlu..HisLeuAspHisGlnAl 76
::: :||| :||| :||| :||| :||| :||| :||| :||| :|||
390 CTGCAGCTACGCAACCGCGACCGCGCGCGCTACACTCTCTACATGAAGG 439

76 aAlaHisGlnProPheProArgProArgPheArgGlnGluThrGlyHisP 93
::: :||| :||| :||| :||| :||| :||| :||| :||| :|||
440 TGGCAA.....GGCGGCC 453

93 roSerLeuGlnArgAspPheProArgSerPheLeu..... 104
::: :||| :||| :||| :||| :||| :||| :||| :||| :|||
454 GTGCCCTGCAGCGCGCGCGCGCGCGCGCGCGAGAGCTTCGACGAGGTGC 503

105 .....LeuAspLeuProAsnPheProAspLeuSerLys..... 115
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
504 CCTCGAGTCCACGCGCACCTACCTGGCGGTGGAGAGCTTCGACGAGGTGC 553

116 AlaAspIleAsnGly...Gln***ProAsnIleGlnValThrIleGluV 131
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
554 TGGCGCTCTCGACCCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAG 603

131 alValAspGlyProAspSerGluAlaAspLysAspGlnHisProGluAsn 147
::: :||| :||| :||| :||| :||| :||| :||| :||| :|||
604 CAGTCTCTGCA.....GATGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 632

148 LysProSerTrpSerValProSerProAspTrpArgAlaTrpTrpGlnAr 164
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
633 GCCCGCACGACGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 682

164 gSerLeu..SerLeuAlaArgAlaAsnSerGlyAspGlnAspTyr***. 179
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
683 ACGACTTCTCGGTGAGTACCTGGTGGTGGGAAACCGCAACCGCGCGCGT 732
```

180 .....TyrAspSerThrSerAspSps 187  
 733 GCCCGCTGCCAGATGCTGCTGCCGCTGGACGCGTGTCTGGCCGGTAG 782  
 187 rAsnPheLeuAsnPro..... 192  
 783 TCGCAGCTCGCACCCCTCGCGGATCATGCAGACCCCTGGCCTGCCTGG 832  
 193 .....ProArgGly... 195  
 833 GCGCGAGGCGGCGGCGGCTGCTGCCGCGGACCCCTGGCCCCCGCGGGGAT 882  
 196 .....TrpAspHisThrAlaProGlyHisArg.....ThrPh 206  
 883 GTCTGCTTGAGAGATGCGGTGGTGGTGGCCCTGAAAACTGCCTCACAG 932  
 206 eGluThrLysAspGlnProGluThrAspSerThrAspGlyGluGlyAspT 223  
 933 CCTGACCCAGGACCGGCGGCGGACGCGGCACAGCGGCGCTGGAAGCTGT 982  
 223 rpSerLeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLys 239  
 983 GGTCCCTGTGGGCGAATGCACGCGGACTCGGGGGAGGCGCTCCAGACG 1032  
 240 ArgThrArgSerCys.....GlyTyrAlaCysTh 249  
 1033 CGGAGCGCACCTGTGCTGCCGCGCGGCGTGGAGGCGGCGCTGCGA 1082  
 249 rAlaThr.....GluSerArgThrCysAspArgProAsnCysProGlyI 264  
 1083 GGGGTGCTGGAGGAGGTCGCCAGTGCACCCGCGGCGCTGC..... 1125  
 264 leGluAspThrPheArgThrAlaAlaThrGluValSerLeuLeuAla... 279  
 1126 ..GGCCCCGCTGGGCGCACAGTCCCGGAGCCAGTCCCTCGCGTCCACA 1173  
 280 .....GlySerGlu.....GluPh 284  
 1174 GATGCCGCGGCGGCGGAGAGTGGGGACGAGCTGCAGCAGTTTGGGTT 1223  
 284 eAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrpM 301  
 1224 CCCAGCCCCCAGACCGGTGACCCAGCAGCCGAGGAGTGTCGCCGTGGA 1273  
 301 etSerCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAsn 317  
 1274 GCGTGTGCTCCAGC..... 1287  
 318 AspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSerTh 334  
 1288 .....ACCTGCGGC..... 1296  
 334 rAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysAspA 351  
 1296 ..... 1296  
 351 laSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTyr 367  
 1297 .....GAGGGCTGGCAGACCCGCGCAGCGCTTC 1323  
 368 CysIleArgSerMetLeuSerLeuGlu...SerThrThrLeuAlaAlaGl 383  
 1324 TCGGTGCTCTCTCTACAGACGAGTGCAGCGGACCCCTCGCGGAGCA 1373  
 383 nHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyA 400  
 1374 GCGGCTGTGC.....AACAACTCTCGGTGCGCCAGTGCATGGTG 1414  
 400 laGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyr 416  
 1415 CCTGG.....GATGAGTGGTGC..... 1431  
 417 LysValAspValLeuProTrpIleIleCys.....LysGl 428

1432 .....CCCTGGAGCCTCTGCTCCAGCACCTGTGGCGCGTGG 1466  
 428 yAspTrpSerArgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysC 445  
 1467 CTTTCGGGATCGACGCGCACCTGTGAGGCCCCCGCCAGTTTGGGGCAACC 1516  
 445 yThrGluSerProSerAspGlu 452  
 1517 CCTGTGAGGGCCCTGAGAAGCAA 1539

OM of: US-09-863-824-2 to: N\_Geneseq\_032802:\* out\_format : pfs  
Date: Apr 25, 2002 10:20 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09863824/runat\_24042002\_150506\_13902/app\_query.fasta\_1.527  
-DB=N\_Geneseq\_032802 -QPMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US09863824.@CGL1\_313 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-863-824-2  
Query length: 464  
Database: N\_Geneseq\_032802:\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 206.440000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2002.DAT:AD23967 +	2513.00	3834.36	3.2e-20		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2002.DAT:AD23969 +	1567.00	2388.05	1.1e-12		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AAH99403 +	1353.00	2062.37	1.6e-11		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2000.DAT:AAH99403 +	1035.00	1574.13	2.5e-79		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AAF93820 +	929.50	1398.73	1.5e-6		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AAF93820 +	929.50	1398.73	1.5e-6		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2002.DAT:AD23968 +	900.00	1371.95	4.5e-68		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA570533 +	836.00	1274.85	1.2e-6		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2000.DAT:AAH99403 +	440.50	662.64	1.5e-28		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570534 +	398.00	602.64	3.2e-2		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570534 +	315.00	472.26	5.9e-1		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA570534 +	232.00	330.17	4.8e-1		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA570534 +	145.50	192.85	0.0215		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	143.50	192.85	0.0215		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	143.00	186.74	0.0470		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	142.00	184.98	0.0589		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	142.00	184.98	0.0589		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	142.00	184.98	0.0589		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	142.00	184.98	0.0589		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	137.50	178.59	0.1338		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	137.50	177.69	0.1501		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	137.50	177.69	0.1501		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	137.50	177.34	0.1570		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	137.00	179.26	0.1328		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA570533 +	133.50	174.87	0.2156		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2002.DAT:AA57179 +	133.50	171.45	0.3341		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA57179 +	133.50	169.14	0.4494		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA57179 +	132.00	175.46	0.1999		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2002.DAT:AA57178 +	132.00	174.18	0.2355		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2000.DAT:AA57178 +	131.50	168.35	0.4976		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA57178 +	131.50	168.07	0.5157		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA57178 +	130.00	170.91	0.3584		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA57178 +	130.00	169.84	0.4107		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA57178 +	130.00	169.84	0.4107		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA57178 +	130.00	169.81	0.4126		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001B.DAT:AA57178 +	129.50	172.68	0.2854		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA57178 +	129.50	172.68	0.2854		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA57178 +	129.50	168.88	0.4649		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA57178 +	129.50	168.88	0.4649		
/SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AA57178 +	128.50	184.75	0.0607		

ID	seq_documentation_block:	seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-embl/NA2002.DAT:AD23967
XX	AD23967 standard; cDNA; 1395 BP.	
AC	AD23967;	
DT	26-MAR-2002 (first entry)	
XX	Human thrombospondin-like protein cDNA #1.	
DE	Human; thrombospondin-like protein; therapeutic; pharmacogenic;	
KW	drug screening; cosmetic; nutraceutical; anti-angiogenic; wound healing;	
KW	endocrine; gene therapy; hyperthyroidism; hypothyroidism; ss.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	1..1395
FT		/*tag= a
FT		/product= "Thrombospondin-like protein"
FT		/transl_except= (pos:364..366, aa:Xaa)
FT		/note= "Xaa is Asn or Val"
FT		/transl_except= (pos:535..537, aa:Xaa)
FT		/note= "Xaa is Lys or Glu"
FT	misc_feature	364
FT		/*tag= b
FT		/note= "This degenerate base represents A-G transition polymorphism"
FT	misc_feature	365
FT		/*tag= c
FT		/note= "This degenerate base represents A-T transversion polymorphism"
FT	misc_feature	535
FT		/*tag= d
FT		/note= "This degenerate base represents A-G transition polymorphism"
PN	WO200190179-A2.	
PD	29-NOV-2001.	
XX		
PF	23-MAY-2001; 2001WO-US16786.	
XX		
PR	23-MAY-2000; 2000US-206415P.	
XX	(LEXI-) LEXICON GENETICS INC.	
XX	Turner CA, Hilbun E, Donoho G, Friedrich G, Abuin A, Zambrowicz B; Sands AT;	
DR	WPI: 2002-089923/12.	
DR	P-PSDB; AAEL4397.	
XX		
PT	New isolated nucleic acid molecule that encodes human thrombospondin-like protein, useful in therapeutic, diagnostic and pharmacogenomic applications, in drug screening and clinical trial monitoring	
PS	Claim 2; Page 35; 39pp; English.	
XX		
CC	The invention relates to nucleic acid molecule encoding human thrombospondin-like protein. The polynucleotide and polypeptide of the invention are useful in therapeutic, diagnostic and pharmacogenomic applications, in drug screening, clinical trial monitoring, and in cosmetic or nutraceutical applications. The polynucleotide is useful for identifying coding sequences and mapping a unique gene to a particular	

CC chromosome, to screen libraries, isolate clones, prepare cloning and  
 CC sequencing templates, as hybridisation probes for screening libraries  
 CC and assessing gene expression patterns, in microarrays and other assay  
 CC formats to screen collections of genetic material from patients who  
 CC have a particular medical condition, to identify mutations associated  
 CC with a particular disease, in a diagnostic or prognostic assay, as  
 CC antisense molecules, and as part of ribozymes and/or triple helix  
 CC sequences that are useful for gene regulation, for detecting mutant  
 CC proteins or inappropriately expressed proteins for the diagnosis of  
 CC disease and in the molecular mutagenesis/revolution of proteins that  
 CC are at least partially encoded by the polynucleotide. The  
 CC thrombospondin-like protein or its modified processed form can  
 CC be used as therapeutics e.g. anti-angiogenic agents, to promote wound  
 CC healing and regulate endocrine functions. The polynucleotide  
 CC can also be used in gene therapy and for treating disorders like  
 CC hyperthyroidism and hypothyroidism. The present sequence is a cDNA  
 CC encoding human thrombospondin-like protein.  
 XX  
 SQ Sequence 1395 BP; 345 A; 431 C; 384 G; 232 T; 3 other;

alignment\_scores:  
 Quality: 2513.00 Length: 464  
 Ratio: 5.439 Gaps: 0  
 Percent Similarity: 99.569 Percent Identity: 99.569

alignment\_block:

US-09-863-824-2 x AAD23967 ..

Align seg 1/1 to: AAD23967 from: 1 to: 1395

1 MetValArgLeuAlaGluLeuLeuLeuLeuGlyLeuLeuLe 17  
 1 ATGTGCGCCTGGCGCGAGCTGCTGCTGGGGCTGCTGCT 50  
 17 uThrLeuHisIleThrValLeuArgGlySerGlyAlaAla 34  
 51 CACGCTGCACATCACCTGCTGCGCGCTGGGAGCCGCGCG 100  
 34 spAlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnLeu 50  
 101 ACGGCGCGCGCGCAAGCCAGCCAGCTGCAGATAACCTCA 150  
 51 ValGlySerAspThrThrSerGluThrSerPheSerLeu 67  
 151 GTGGAGTGCACACCATCAAGAACCAAGCTTTCTCTCTC 200  
 67 proArgGluHisLeuAspHisGlnAlaAlaHisGlnPro 84  
 201 ACCAAGGAGCATCTGGACCAACAGCTGCACCAACCTTC 250  
 84 roArgPheArgGlnGluThrGlyHisProSerLeuGlnA 100  
 251 CGCGATTCCGACAGAGCGGCGACCTTCATTGCAAGA 300  
 101 ArgSerPheLeuLeuAspLeuProAsnPheProAspLe 117  
 301 AGATCTTCTCTGATCTACCAACATTCAGATCTTCCAA 350  
 117 pIleAsnGlyGln\*\*ProAsnIleGlnValThrIleGlu 134  
 351 TATCAATGGCGAGRTCCCAATATCCAGTCCACATAG 400  
 134 lyProAspSerGluAlaAspLysAspGlnHisProGlu 150  
 401 GTCCCTGACTCTGAAGCAGATAAGATCAGCATCCGGA 450  
 151 TrpSerValProSerProAspTrpArgAlaTrpTrpG 167  
 451 TGTTCAGTCCCATCCCGACTGGCGGCTGGTGGCAGAG 500  
 167 rLeuAlaArgAlaAsnSerGlyAspGlnAspTyr\*\*Ty 184

501 CTTGGCCAGGGCAACAGCGGGACAGGACTACRAGTACG 550  
 184 erAspAspSerAsnPheLeuAsnProArgGlyTrpAsp 200  
 551 CAGACGACAGCAACTCTCTCAACCCCGGGGTGGACCA 600  
 201 ProGlyHisArgThrPheGluThrLysAspGlnProG 217  
 601 CCAGGCCACCGGACTTTTCAAAACCAAGATCAGCCAG 650  
 217 rAspGlyGluGlyAspTrpSerLeuTrpSerValThr 234  
 651 AGATGCGAGGCTGACTGGAGTCTCTGGTCTGTCTCA 700  
 234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyT 250  
 701 GGNACGGCAACCAAGAACCGGACCGGCTCTGTGGCT 750  
 251 ThrGluSerArgThrCysAspArgProAsnCysProG 267  
 751 ACAGAAATCGAGGACCTGTGACCGTCCAACTGCCCA 800  
 267 rPheArgThrAlaAlaThrGluValSerLeuLeuAla 284  
 801 TTTAGGACAGCTGCCACCAAGTGTCTGTGGGGAAGC 850  
 284 heAsnAlaThrLysLeuPheGluValAspThrAspS 300  
 851 TTAATGCCACCAACTGTTTGAAGTTGACACAGACAG 900  
 301 MetSerCysLysSerGluPheLeuLysTyrMetHis 317  
 901 ATGAGCTGCAAAAGCGAGTTCTTAAAGAGTATGCA 950  
 317 nAspLeuProSerCysProCysSerTyrProThrGlu 334  
 951 TGACCTGCCAGCTGCCCTGCTCTACCCACTGAGTG 1000  
 334 hrAlaAspIlePheAspArgIleLysArgLysAsp 350  
 1001 CGCCCGACATCTTCGACCGCATCAAGCGCAAGGAC 1050  
 351 AlaSerGlyProLysGluLysLeuGluIleTyrLys 367  
 1051 GCACGGCGGCCCAAGAGAGCTGGAGATCTACAAGC 1100  
 367 rCysIleArgSerMetLeuSerLeuGluSerThrTh 384  
 1101 CTGCATCCGCTCCATGCTGCTCCCTGGAGACCA 1150  
 384 isCysCysTyrGlyAspAsnMetGlnLeuIleThr 400  
 1151 ACTGCTGTACGGCGACACATGACGCTCATCACCAG 1200  
 401 GlyThrProAsnLeuIleSerThrGluPheSerAla 417  
 1201 GGCACGCCCAACCTCATCAGCACCGAGTTCTCCG 1250  
 417 svalaspValLeuProTrpIleIleCysLysGlyAs 434  
 1251 GGTGGAGCTCTGCTCCCTGGATTATCTCAAGGGT 1300  
 434 snGluAlaArgProProAsnAsnGlyGlnLysCys 450  
 1301 ACAGAGCCCGCTCCCAACACCGACAGAGTGCACAG 1350  
 451 AspGluAspTyrIleLysGlnPheGlnGluAlaArg 464  
 1351 GACGAGGACTACATCAAGCAGTTTCAAGAGCCAGG 1392

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD23969

seq\_documentation\_block:



267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerCluGluP 284  
|||||  
801 TTTTAGGACAGCTCCACCGAGTGTCTGCTGGCGGAGGAGGAGT 850  
|||||  
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCys 297  
|||||  
851 TTAATGCCACCAACTGTTGAAGTTGCTCCAGCATGT 891  
|||||

seq\_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/WA2001A.DAT:AAH99403

267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerCluGluP 284  
|||||  
801 TTTTAGGACAGCTCCACCGAGTGTCTGCTGGCGGAGGAGGAGT 850  
|||||  
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCys 297  
|||||  
851 TTAATGCCACCAACTGTTGAAGTTGCTCCAGCATGT 891  
|||||

seq\_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/WA2001A.DAT:AAH99403

seq\_documentation\_block:  
ID AAH99403 standard; cDNA; 739 BP.  
AC AAH99403;  
XX  
XX 16-OCT-2001 (first entry)  
XX Human protein encoding cDNA sequence SEQ ID NO:238.  
XX  
XX Human; cancer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnerary; antilucer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200153455-A2.  
XX  
XX 26-JUL-2001.  
XX  
XX 22-DEC-2000; 2000WO-US35017.  
XX  
XX 23-DEC-1999; 99US-0471275.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX  
XX 25-APR-2000; 2000US-0552317.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-457603/49.  
XX  
XX P-PSDB; AAM25462.  
XX  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
XX Claim 1; Page 403-404; 1217pp; English.  
XX  
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX AAM25963. The proteins can have activities based on the tissues and  
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;  
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;  
XX antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;  
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX encoding them can be used in gene therapy, antisense therapy and vaccine  
XX production. The proteins and polynucleotides are useful for screening for  
XX agonists or antagonists of a protein and for the treatment and diagnosis  
XX of disorders associated with the activity of a protein e.g. inflammation,  
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

seq\_documentation\_block:  
ID AAH99403 standard; cDNA; 739 BP.  
AC AAH99403;  
XX  
XX 16-OCT-2001 (first entry)  
XX Human protein encoding cDNA sequence SEQ ID NO:238.  
XX  
XX Human; cancer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnerary; antilucer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200153455-A2.  
XX  
XX 26-JUL-2001.  
XX  
XX 22-DEC-2000; 2000WO-US35017.  
XX  
XX 23-DEC-1999; 99US-0471275.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX  
XX 25-APR-2000; 2000US-0552317.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-457603/49.  
XX  
XX P-PSDB; AAM25462.  
XX  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
XX Claim 1; Page 403-404; 1217pp; English.  
XX  
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX AAM25963. The proteins can have activities based on the tissues and  
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;  
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;  
XX antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;  
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX encoding them can be used in gene therapy, antisense therapy and vaccine  
XX production. The proteins and polynucleotides are useful for screening for  
XX agonists or antagonists of a protein and for the treatment and diagnosis  
XX of disorders associated with the activity of a protein e.g. inflammation,  
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers, allergic  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX  
XX Sequence 739 BP; 187 A; 225 C; 206 G; 121 T; 0 other;

alignment\_scores:  
Quality: 1353.00 Length: 246  
Ratio: 5.522 Gaps: 0  
Percent Similarity: 99.593 Percent Identity: 99.593  
alignment\_block:  
US-09-863-824-2 x AAH99403  
Align seg 1/1 to: AAH99403 from: 1 to: 739  
175 AspGlnAspTyr\*\*\*TyrAspSerThrSerAspAspSerAsnPheLeuAs 191  
|||||  
2 GACCAGGACTACAAGTACGACAGTACCTCAGCAGCAGCACTTCCTCAA 51  
191 nProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGluT 208  
|||||  
52 CCCCCCAGGGGGTGGGACCATACAGCCCGCAGGCGACCGGACTTTTGA 101  
208 hrLysAspGlnProGluTyrAspSerThrAspGlyGluGlyAspTrpSer 224  
|||||  
102 CCAAAGATCACGCCAGATATGATTCACAGATGCGAGGCTGACTGGAGT 151  
225 LeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysArgTh 241  
|||||  
152 CTCTGGTCTGTCTGCAGCGTCACCTCGGGGAACGGCAACCAAGAAACGGAC 201  
241 rArgSerCysGlyTyrAlaCysThrAlaThrGluSerArgThrCysAspA 258  
|||||  
202 CCGGTCTTGTGGCTACGCGTCACTGCAACAGAAATCGAGGACCTGTGACC 251  
258 rGProAsnCysProGlyIleGluAspThrPheArgThrAlaAlaThrGlu 274  
|||||  
252 GTCCAAACTGCCAGGAATTGAAGACACTTTTAGGACAGTGCCACCGAA 301  
275 ValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheGl 291  
|||||  
302 GTGAGTCTGCTTGGCGGAAGCGAGAGTTAATGCCACCAAACTGTTTGA 351  
291 uValAspThrAspSerCysGluArgTTPMetSerCysLysSerGluPheL 308  
|||||  
352 AGTTGACACAGACAGCTGTGAGCGCTGATGAGCTGCCAGCTGCCCTGCT 401  
308 euLysLysTyrMetHisLysValMetAsnAspLeuProSerCysProCys 324  
|||||  
402 TAAAGAAGTACATGCACAAAGGTGATGAATGACCTGCCAGCTGCCCTGTC 451  
325 SerTyrProThrGluValAlaTyrSerThrAlaAspIlePheAspArgIl 341  
|||||  
452 TCCTACCCCACTAGGTGGCTTACAGACGCGCGGACATCTTCGACCGCAT 501  
341 eLysArgLysAspPheArgTTPAspAlaSerGlyProLysGluLysL 358  
|||||  
502 CAAGCGCAAGGACTTCCGCTGGAAGAGCGCCAGCGGGGCCAAGGAGAAGC 551  
358 euGluIleTyrLysProThrAlaArgTyrCysIleArgSerMetLeuSer 374  
|||||  
552 TGGAGATCTACAAGCCCACTGCCCGGTACTGTCATCCGCTCCATGCTGCC 601  
375 LeuGluSerThrThrLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMe 391  
|||||  
602 CTGGAGAGACCAACGCTGGCGGACAGCAGCACTGCTGCTACGGCGCAACAT 651





ID AAF93820 standard; cDNA; 2981 BP.  
 AC AAF93820;  
 XX 23-MAY-2001 (first entry)  
 DT  
 XX Human cDNA encoding a membrane or secretory protein clone PSEC0137.  
 DE  
 XX Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX EP1067182-A2.  
 PN  
 XX 10-JAN-2001.  
 PD  
 XX 07-JUL-2000; 2000EP-0114090.  
 PF  
 XX 08-JUL-1999; 99JP-0194179.  
 XX 11-JAN-2000; 2000JP-0118775.  
 PR  
 XX 02-MAY-2000; 2000JP-0183766.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 PI WPI; 2001-093989/11.  
 XX P-PSDB; AAB88393.  
 DR  
 XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 XX gene therapy or as candidate target molecules in drug development -  
 PT  
 XX Claim 1; SEQ ID 153; 609pp + CD ROM; English.  
 PS  
 XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by  
 CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic  
 CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
 CC (ELISA). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.  
 XX  
 SQ Sequence 2981 BP; 609 A; 925 C; 887 G; 560 T; 0 other;

alignment\_scores:  
 Quality: 929.50 Length: 500  
 Ratio: 2.923 Gaps: 18  
 Percent Similarity: 63.600 Percent Identity: 41.600

alignment\_block:  
 US-09-863-824-2 x AAF93820 ..  
 Align seg 1/1 to: AAF93820 from: 1 to: 2981

25 ArgGlySerGlyAlaAlaAspGlyProAspAlaAlaAlaGlyAsnAlaSe 41  
 ||||| : : : : : ||| ||||| : : : : :  
 357 AGAGGTACTCGTTCGGCTGGAGCTGCAGAGCTGCCGGGATTCGC... 404  
 41 rGlnAlaGlnLeuGlnAsnAsnLeuVal..... 51  
 ||||| : : : : : ||||| : : : : :  
 405 .....CAGCACAACTTGAGTACCCCTAACCTGATACCCAG 441  
 52 .....GlySerAspThr 55  
 ||||| : : : : : ||||| : : : : :  
 442 GCTTCAGCCTCCAGATCCTAGGCCTCTGAGGGAAGAGGAGGAGGCACG 491  
 56 ThrSerGluThrSerPheSerLeuSerLysGluAlaProArgGluHisLe 72  
 ||||| : : : : : ||||| : : : : :  
 492 ACTGCTCCCAAGAACCCAGCTGAGGAGGAGTACACACATGGATGT 541  
 72 uAspHisGlnAlaAlaHisGlnProPheProArgProArgPheArgGlnG 89  
 ||||| : : : : : ||||| : : : : :  
 542 GGACTGTCACTGAGCCAGCAGCCCTGACCCAGG.....G 576  
 89 lu.ThrGlyHisProSerLeuGlnArgAspPheProArgSerPheLeuLe 105  
 ||||| : : : : : ||||| : : : : :  
 577 AATGCCAGCCTCCAGGAGCCAGGAGGTACTCC.....TTGCTGCT 620  
 105 uAspLeuProAsnProAspLeuSerLysAlaAspIleAsnGlyGln\* 122  
 ||||| : : : : : ||||| : : : : :  
 621 GGAGCTCAGAAAGCTGCCAATTTGGTCCACGCAACCTTGATACCCCTA 670  
 122 \*\*ProAsnIleGlnValThrIleGluValValAspGlyProAspSerGlu 138  
 ||||| : : : : : ||||| : : : : :  
 671 ACCTGTATACCAAGTACCACTCAAGTGTGGAGGAGCCAGGCGGAG 720  
 139 AlaAspLysAsp.....GlnHisProGluAsnLysPro..... 149  
 ||||| : : : : : ||||| : : : : :  
 721 GTGTGATAGACCTGTGTGGTGTAGCCAGCAATCCCGCCAGGATAC 770  
 150 ....SerTrpSerValProAspTrpArgAlaTrp..... 162  
 ||||| : : : : : ||||| : : : : :  
 771 CCTTATGCTGG...CTGCCCGCCCTC...TGGCCCTCTCTGGGGAGACT 814  
 163 .....GlnArgSerLeuSerLeuAlaArgAlaAsnSerGly..... 174  
 ||||| : : : : : ||||| : : : : :  
 815 ACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 864  
 175 .....AspGlnAspTyr..... 178  
 ||||| : : : : : ||||| : : : : :  
 865 GAGGAGACGAGGAGTATCTTCAGAGGATATCAGGGGTGAGGATCAAGA 914  
 179 .....\*\*\*TyrAspSerThrS 184  
 ||||| : : : : : ||||| : : : : :  
 915 GCACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 964  
 184 eAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAla 200  
 ||||| : : : : : ||||| : : : : :  
 965 CAGAG.....AACTGGGACAGGAGGCTGG.....CTGGCC 993  
 201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217  
 ||||| : : : : : ||||| : : : : :  
 994 CCGGGGATTTGGTCTTC.....AAGATTCTGTACGTACGAC...TA 1034  
 217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234  
 ||||| : : : : : ||||| : : : : :  
 1035 TCAGCCTCAGAAGGAGTGGAGTCCCTGCTCCCTGCAGTGGGAAGTCA 1084  
 234 lYAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250  
 ||||| : : : : : ||||| : : : : :  
 1085 GCATGGCAAGCAGCAGGAGTCCGCTGTGGCTATGGCTGCTGCTGCTG 1134  
 251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGlu..... 265  
 ||||| : : : : : ||||| : : : : :  
 1135 ACCGAGACCCGTACCTGTGACCTGCCTCTCTGCTGCTGCTGCTGCTG 1184  
 266 AspThrPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerG 282

```

|||||... 1185 GGACACCTTGGCGCTCCCACTGAGGAGTGAAGCTCCTGGCCGCG... 1230
282 luGluPheAsnAlaThrLysLeuPheGluValAspThrAspSerCysGlu 298
1231 .....AATGCTACGGACATGATCAAGATGTGGACAGCTGTGAG 1272
299 ArgTrpMetSerCysLysSerGluPheLeuLysLysTyrMetHisLysVa 315
1273 AAGTGGCTGAACCTCAAGAGCGACTTCTTAATCAAGTATCTGAGCCAGAT 1322
315 lMetAsnAspLeuProSerCysProCysSerTyrProThrGluValAla 332
1323 GCTCGGAGACCTGCCAGCTGCCGCTGTGCTACCCACTGGAGGCCATGG 1372
332 yrSerThrAlaAspIlePheAspGlyLysArgLysAspPheArgTrp 348
1373 ACAGCCCTGTGAGCTACAGACGAGCAGACCGGCCGCGAGCTTCGGGTGG 1422
349 LysAspAlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAl 365
1423 AGGATGCCAGTGGCCCTCGCGAGCGCTGGACATCTACCAAGCCACGCG 1472
365 aArgTyrCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAla 382
1473 GCGCTTCTGCTGCTGCTTCATGCTGTCTGGGGAGAGCAGCACACTGGCG 1522
382 laGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLys 398
1523 CCCAGCACTGCTGTATGACGAGCAGACCGCGCTGCTGACCCGTGGCAAG 1572
399 GlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHi 415
1573 GCGCGCGCATGCCCAACCTCATCAGCACCGCACTTCTCACCTAAGCTGCA 1622
415 sTyrLysValAspValLeuProTrpIleIleCysLysGlyAspTrpSerA 432
1623 CTTCAAGTTCGACACGACGCGCTGATGCTGTGCAAGGGGAGCTGGAGCC 1672
432 rgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSer 448
1673 GCCTCCAGCGTGTCTCCTCCCAACACAGCGCGGAGCTGCACCGCAAC 1722
449 ProSerAspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464
1723 CCCCTGGAGGAGGAGTACCTAGCACAGTTCGAGGAGGCCAAGGAGTAC 1770

```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:AAF29348

seq\_documentation\_block:

ID AAF29348 standard; cDNA; 2981 BP.

XX AC AAF29348;

XX XX 20-APR-2001 (first entry)

XX DE Human proliferation differentiation factor cDNA sequence.

XX KW Human; proliferation differentiation factor; haematopoietic function; ss.

XX OS Homo sapiens.

XX XX WO200104312-A1.

XX PD 18-JAN-2001.

XX XX 06-JUL-2000; 2000WO-JP04514.

XX PF 08-JUL-1999; 99JP-0194179.

XX PR 18-OCT-1999; 99US-0159586.

XX XX (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;  
 XX WPI: 2001-138354/14.  
 DR P-PSDB; AAB49765.  
 XX Proliferation-differentiation factor protein encoded by PSEC137 cloned  
 PT from human cDNA library, being hematopoietic factor for inducing  
 PT differentiation of blood cells, used to maintain hematopoietic function  
 XX  
 XX  
 PS Claim 1; Page 36-41; 49pp; Japanese.  
 XX This invention relates to polynucleotide sequence PSEC137, which encodes  
 CC a human proliferation differentiation factor protein. Included in the  
 CC invention is a vector containing the cDNA sequence, a transformant  
 CC containing the vector, and a process for producing the protein. The  
 CC protein together with its encoded DNA and antisense DNA are used in drugs  
 CC to maintain the haematopoietic function. The present sequence represents  
 CC cDNA encoding the proliferation differentiation factor protein.  
 XX  
 SQ Sequence 2981 BP; 609 A; 925 C; 887 G; 560 T; 0 other;

alignment\_scores:

Quality: 929.50 Length: 500  
 Ratio: 2.923 Gaps: 18  
 Percent Similarity: 63.600 Percent Identity: 41.600

alignment\_block:

US-09-863-824-2 x AAF29348 ..

Align seg 1/1 to: AAF29348 from: 1 to: 2981

```

25 ArgGlySerGlyAlaAlaAspGlyProAspAlaAlaGlyAsnAlaSe 41
||||| : : : : : ||| ||||| : : :
357 AGAGGTACTCCGTTCGGTGGCTGAGCTGCAGAGCTGCCGGGATGGC... 404
41 rGlnAlaGlnLeuGlnAsnAsnLeuAsnVal..... 51
||||| : : : : :
405 .....CAGCACACCTTGAGTACCCCTAACCTGATACCCAG 441
52 .....GlySerAspThr 55
||||| : : : : :
442 GCTTCAGCCTCCCGATCCTAGGCCTCTGAGGGAAGAGAGGAGGCACG 491
56 ThrSerGluThrSerPheSerLeuSerLysGluAlaProArgGluHisLe 72
||||| : : : : :
492 ACTGCTCCCCAGAACCCACCTGTCAGGCAGAGCTACACCAACATGGATGT 541
72 uAspHisGlnAlaAlaHisGlnProPheProArgProArgPheArgGlnG 89
||||| : : : : :
542 GGACTGTCACTGAGCCAGCAGCCCTGACCCCGG.....G 576
89 lu.ThrGlyHisProSerLeuGlnArgAspPheProArgSerPheLeuLe 105
|| : : : : :
577 AATGCCAGCCTCCCGAGGACCCAGGAGGTACTCC.....TTGCTGCT 620
105 uAspLeuProAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln* 122
||||| : : : : :
621 GGAGCTGCAGAAAGTGCAGAAATTGGTCCACGCAACCTTGATGATACCCCTA 670
122 **ProAsnIleGlnValThrIleGluValValAspGlyProAspSerGlu 138
||||| : : : : :
671 ACCCTGATACCAAGTGACCATCAAGGTGGTGGAGGACCCCGCCGCGAG 720
139 AlaAspLysAsp.....GlnHisProGluAsnLysPro..... 149
: : : : :
721 GTGTCGATAGACCTGTGGCTGAGCCAGCAATCCCCCGCCCGAGGATAC 770
150 ....SerTrpSerValProSerProAspTrpArgAlaTrpTrp..... 162
||||| : : : : :
771 CCTTAGCTGG...TGCCCGCCCTC...TGCCCTTCTCTGGGAGACT 814

```



CC hyperthyroidism and hypothyroidism. The present sequence is a cDNA  
 CC encoding human thrombospondin-like protein.  
 XX  
 SQ Sequence 495 BP; 126 A; 156 C; 139 G; 74 T; 0 other;

alignment\_scores:  
 Quality: 900.00 Length: 164  
 Ratio: 5.488 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-863-824-2 x AAD23968 ..

Align seg 1/1 to: AAD23968 from: 1 to: 495

301 MetSerCysLysSerGluPheLeuLysTyrMetHisLysValMetAs 317  
 |||||  
 1 ATGAGCTGCAAAAGCGAGTCTTAAAGAAAGTACATGCACAAGGTGATGAA 50  
 317 nAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSert 334  
 |||||  
 51 TGACCTGCCAGCTGCCTCTCTACCCCACTGAGGTGCCCTACAGCA 100  
 334 hrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysasp 350  
 |||||  
 101 CGGCGGACATCTTCAGCGCATCAAGCGCAAGGACTTCGCTGGAAGGAC 150  
 351 AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367  
 |||||  
 151 GCCAGCGGCGCAAGGAGAGCTGGAGATCTACAAGCCACCTGCCCGGTA 200  
 367 rCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGlnH 384  
 |||||  
 201 CTGATCCGCTCCATGCTCTCTGGAGAGCACCACGCTGGCGGCACAGC 250  
 384 IsCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAla 400  
 |||||  
 251 ACTGCTGTACGGGACAAACATGCAGCTCATCACCAGGGGCAAGGGGCG 300  
 401 GlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyrLy 417  
 |||||  
 301 GGCAGCGCCCAACCTCATCAGCAGCCAGGTCTCCGCGGAGCTCCACTACA 350  
 417 sValAspValLeuProTyrPheIleCysLysLysGlyAspTrpSerArgTyrA 434  
 |||||  
 351 GGTGGACGCTCTGCCTGGATTTACTGCAAGGGTGACTGGAGCAGGTATA 400  
 434 snGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSerProSer 450  
 |||||  
 401 ACGAGCGCGGCTCCCAACAACGACAGAGTGACACAGAGAGCCCTCG 450  
 451 AspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464  
 |||||  
 451 GACGAGGACTACATCAAGCAGTCTCAAGAGGCGCAGGAATAT 492

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001B.DAT.AAS70533

seq\_documentation\_block:

ID AAS70533 standard; cDNA; 448 BP.

XX

AC AAS70533;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #6337.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; ABG06346.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 1; SEQ ID No 6337; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 448 BP; 123 A; 122 C; 123 G; 80 T; 0 other;

alignment\_scores:

Quality: 836.00 Length: 149  
 Ratio: 5.649 Gaps: 0  
 Percent Similarity: 99.329 Percent Identity: 99.329

alignment\_block:

US-09-863-824-2 x AAS70533 ..

Align seg 1/1 to: AAS70533 from: 1 to: 448

175 AspGlnAspTyr\*\*\*TyrAspSerThrSerAspSerAspPheLeuAs 191  
 |||||  
 2 GACCAGGACTACAAGTACGACAGTACCTCAGACGACAGCAACTCTCTCAA 51  
 191 nProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGluT 208  
 |||||  
 52 CCCCCCAGGGGTGGGACCACATACAGCCCGGACCCGAGCTTTTGAA 101  
 208 hrLysAspGlnProGluTyrAspSerThrAspGlyGluGlyAspTrpSer 224  
 |||||  
 102 CCAAGATCAGCCAGCAATATGATTCACAGATGGCGAGGGTCACTGGAGT 151  
 225 LeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysArgTh 241  
 |||||  
 152 CTCTGGTCTGTCTGCAGCGTCACCTGCGGGAACGCAACCAACGACGAC 201

241 rArgSerCysGlyTyrAlaCysThrAlaThrGluSerArgThrCysAspA 258  
 |||||  
 202 CCGGTCCTGGTACGGTGCACATGCAACAGATCGAGGACCTGTGACC 251  
 |||||  
 258 rgProAsnCysProGlyTleGluAspThrPheArgThrAlaAlaThrGlu 274  
 |||||  
 252 GTCCAAACTGCCAGGAATTAAGACACACTTTTAGGACAGCTGCCACCGAA 301  
 |||||  
 275 ValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheG1 291  
 |||||  
 302 GTGAGTCGTCTGCGGGAAGGAGGATTTAATGCCACCAACTGTGTGA 351  
 |||||  
 291 uValAspThrAspSerCysGluArGTrpMetSerCysLysSerGluPheL 308  
 |||||  
 352 AGTTGCACACAGACAGCTGTGAGCGCTGTGATGAGCTGCAAAAGCGAGTCT 401  
 |||||  
 308 euLysLysTyrMethIleLysValMetAsnAspLeuProSerCysPro 323  
 |||||  
 402 TAAAGAAGTACATGCACAAAGGTGATGATGACCTGCCCGAGCTGCCCC 448  
 |||||

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAC76131

seq\_documentation\_block:  
 ID AAC76131 standard; cdna; 790 BP.

XX AAC76131;  
 AC  
 DT 08-FEB-2001 (first entry)  
 DE Human ORFX ORF1686 polynucleotide sequence SEQ ID NO:3371.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.  
 OS  
 PN WO200058473-A2.

XX 05-OCT-2000.  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.  
 XX Shinkets RA, Leach M;  
 PI WPI; 2000-602362/57.  
 DR P-PSDB; AAB41922.

XX Novel nucleic acids and peptides derived from open reading frame x,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX Claim 5; Page 2553; 5507pp; English.  
 PS AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 XX

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 XX Sequence 790 BP; 181 A; 238 C; 212 G; 157 T; 2 other;

alignment\_scores:  
 Quality: 440.50 Length: 91  
 Ratio: 5.122 Gaps: 1  
 Percent Similarity: 94.505 Percent Identity: 91.209

alignment\_block:  
 US-09-863-824-2 x AAC76131/rev ..

Align seg 1/1 to reverse of: AAC76131 from: 1 to: 790  
 160 AlaTrpTrpGlnArg.....SerLeuSerLeuAlaArgAlaAsnSe 173  
 |||||  
 275 GCCTGGTGGTCCAGATGCTCCCTGGTGTCTTTGGCAGGGCAAA.AG 227  
 |||||  
 173 rGlyAspGlnAspTyr\*\*\*TyrAspSerThrSerAspSerAsnPhel 190  
 |||||  
 226 CGGGACACAGACTACAGTACGGCAGTACCTACAGACAGACAGCACTCC 177  
 |||||  
 190 euAsnProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPhe 206  
 |||||  
 176 TCACCCCCCAGGGGGTGGGACCATACAGCCCGCCAGCCCGGACTTT 127  
 |||||  
 207 GluThrLysAspGlnProGluTyrAspSerThrAspGlyGluGlyAspTr 223  
 |||||  
 126 GAAACCAAGATCAGCCAGAAATATGATTCCACAGATGGCGAGGGTGACTG 77  
 |||||  
 223 pSerLeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysA 240  
 |||||  
 76 GAGTCTCTGGTCTGTCTGACGGTCACTCGGGGAACGGCAACCAAGAAC 27  
 |||||  
 240 rgThrArgSerCysGlyTyrAla 247  
 |||||  
 26 GGACCCGGTCTGTGGCTACGGG 4

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS34028

seq\_documentation\_block:  
 ID AAS34028 standard; cdna; 478 BP.

XX AAS34028;  
 AC  
 DT 17-DEC-2001 (first entry)

XX Human cdna encoding a novel foetal antigen, SEQ ID No 552.  
 DE Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;  
 KW immunomodulator; cardiovascular; cytostatic; nephroprotective;  
 KW cardiovascular; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; breast neoplasm; cancer;  
 KW

KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angioneurosis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.

OS Homo sapiens.

XX WO200155312-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01321.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225477.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226868.

XX 22-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227009.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229343.

XX 01-SEP-2000; 2000US-0229344.

XX 01-SEP-2000; 2000US-0229345.

XX 05-SEP-2000; 2000US-0229509.

XX 05-SEP-2000; 2000US-0229513.

XX 06-SEP-2000; 2000US-0230437.

XX 06-SEP-2000; 2000US-0230438.

XX 08-SEP-2000; 2000US-0231242.

XX 08-SEP-2000; 2000US-0231243.

XX 08-SEP-2000; 2000US-0231244.

XX 08-SEP-2000; 2000US-0231413.

XX 08-SEP-2000; 2000US-0231414.

XX 08-SEP-2000; 2000US-0232080.

XX 08-SEP-2000; 2000US-0232081.

XX 12-SEP-2000; 2000US-0231968.

XX 14-SEP-2000; 2000US-0232397.

XX 14-SEP-2000; 2000US-0232398.

XX 14-SEP-2000; 2000US-0232399.

XX 14-SEP-2000; 2000US-0232400.

XX 14-SEP-2000; 2000US-0232401.

XX 14-SEP-2000; 2000US-0233063.

XX 14-SEP-2000; 2000US-0233064.

XX 14-SEP-2000; 2000US-0233065.

XX 21-SEP-2000; 2000US-0234223.

XX 21-SEP-2000; 2000US-0234274.

XX 23-SEP-2000; 2000US-0234997.

XX 23-SEP-2000; 2000US-0234998.

XX 26-SEP-2000; 2000US-0235484.

XX 27-SEP-2000; 2000US-0235834.

XX 27-SEP-2000; 2000US-0235836.

XX 29-SEP-2000; 2000US-0236327.

XX 29-SEP-2000; 2000US-0236367.

XX 29-SEP-2000; 2000US-0236368.

XX 29-SEP-2000; 2000US-0236369.

XX 29-SEP-2000; 2000US-0236370.

XX 02-OCT-2000; 2000US-0236802.

XX 02-OCT-2000; 2000US-0237037.

XX 02-OCT-2000; 2000US-0237038.

XX 02-OCT-2000; 2000US-0237039.

XX 02-OCT-2000; 2000US-0237040.

XX 13-OCT-2000; 2000US-0239935.

XX 13-OCT-2000; 2000US-0239937.

XX 20-OCT-2000; 2000US-0240960.

XX 20-OCT-2000; 2000US-0241221.

XX 20-OCT-2000; 2000US-0241785.

XX 20-OCT-2000; 2000US-0241786.

XX 20-OCT-2000; 2000US-0241787.

XX 20-OCT-2000; 2000US-0241808.

XX 20-OCT-2000; 2000US-0241809.

XX 20-OCT-2000; 2000US-0241826.

XX 01-NOV-2000; 2000US-0244617.

XX 08-NOV-2000; 2000US-0246474.

XX 08-NOV-2000; 2000US-0246475.

XX 08-NOV-2000; 2000US-0246476.

XX 08-NOV-2000; 2000US-0246477.

XX 08-NOV-2000; 2000US-0246478.

XX 08-NOV-2000; 2000US-0246523.

XX 08-NOV-2000; 2000US-0246524.

XX 08-NOV-2000; 2000US-0246525.

XX 08-NOV-2000; 2000US-0246526.

XX 08-NOV-2000; 2000US-0246527.

XX 08-NOV-2000; 2000US-0246528.

XX 08-NOV-2000; 2000US-0246532.

XX 08-NOV-2000; 2000US-0246609.

XX 08-NOV-2000; 2000US-0246610.

XX 08-NOV-2000; 2000US-0246611.

XX 08-NOV-2000; 2000US-0246613.

XX 17-NOV-2000; 2000US-0249207.

XX 17-NOV-2000; 2000US-0249208.

XX 17-NOV-2000; 2000US-0249209.

XX 17-NOV-2000; 2000US-0249210.

XX 17-NOV-2000; 2000US-0249211.

XX 17-NOV-2000; 2000US-0249212.

XX 17-NOV-2000; 2000US-0249213.

XX 17-NOV-2000; 2000US-0249214.

XX 17-NOV-2000; 2000US-0249215.

XX 17-NOV-2000; 2000US-0249216.

XX 17-NOV-2000; 2000US-0249217.

XX 17-NOV-2000; 2000US-0249218.

XX 17-NOV-2000; 2000US-0249244.

XX 17-NOV-2000; 2000US-0249245.

XX 17-NOV-2000; 2000US-0249264.

XX 17-NOV-2000; 2000US-0249265.

XX 17-NOV-2000; 2000US-0249297.

XX 17-NOV-2000; 2000US-0249299.

XX 17-NOV-2000; 2000US-0249300.

XX 01-DEC-2000; 2000US-0250160.

XX 01-DEC-2000; 2000US-0250391.

XX 05-DEC-2000; 2000US-0251030.

XX 05-DEC-2000; 2000US-0251988.

XX 05-DEC-2000; 2000US-0256719.

XX 06-DEC-2000; 2000US-0251479.

XX 08-DEC-2000; 2000US-0251856.

XX 08-DEC-2000; 2000US-0251868.

XX 08-DEC-2000; 2000US-0251869.

XX 08-DEC-2000; 2000US-0251989.











alignment_scores:		Quality: 145.50	Length: 328
		Ratio: 0.951	Gaps: 17
		Percent Similarity: 46.646	Percent Identity: 24.085
alignment_block:		US-09-863-824-2 x AAC90311	
Align seg 1/1 to: AAC90311		from: 1 to: 4339	
15	LeuLeuLeuLeuLeuHisIleThrValLeuArgGlySerGlyAlaAla.A	31	
923	GTCCCTCTTACCTTGACAAACAGCTGTGAACGGTTCACGCCCTGTCTAT	972	
31	spGlyProAspAlaAlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsn	47	
973	CGCCACCACTACATCGGCC.....ACAAACAAAGGACCTCC	1010	
48	AsnLeuAsnValGlySerAspThrThrSerGluThrSerPheSerLeu.S	64	
1011	AAGCTATCTGTGGCTCTCTCTGTGATGAACATCCAGCATGGTCTCGAA	1060	
64	erLysGluAlaProArgGluHisLeuAspHisGlnAlaAlaHisGlnPro	80	
1061	CT...GAAGGGCTCGGCACCATCTGTACCACCTCTGCAGGACGATCCG	1107	
81	PheProArgProArgPheArgGln.....GluThrGlyHi	92	
1108	AAAGTGACGGAAGAGACAGAGAGCTGGTCACTAGCTGAAGCGGCCTC	1157	
92	sProSerLeuGlnArg.....Aspp	99	
1158	CCCTCTGCTTTACATGAGTCCAGTACAAACAGCAGAGTGACT	1207	
99	heProArgSerPheLeuLeuAspLeuPro.AsnPheProAspLeuSerLy	115	
1208	GTACACAGTTGCACAGAGTGTCACTGCCAAGACTCGGTTACCATCTGCAA	1257	
115	sAlaAspIleAsnGlyGln**ProAsnIleGlnValThrIleGluVal	132	
1258	AAAGGTGCTCTGCCCATCATGCCCTGCTCCACGCCACAGTT.....	1300	
132	alAspGlyProAspSerGluAlaAspLysAspGlnHisProGluAsnLys	148	
1301	.....CCTGATGCTGAATGC.....TGCCACGGTCTCTGG	1330	
149	ProSerTrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSe	165	
1331	CCACGCGACTCTGCT..GACGATGGTGTCTCCCTGTCTGAGTGGAC	1377	
165	rLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAsp.....Tyr**T	180	
1378	CTCCTGCTGCCACATGTGGCAATGGAATTAGCAACGTGGTCTCTCT	1427	
180	yrAspSerThrSerAsp.....	185	
1428	GTGACAGCCTCAACACAGATGCGAGGGCTCTTCGGTACAGCAGGAGCC	1477	
186	.....AspSerAsnPheLeuAsnProArgGlyTr	196	
1478	TGCCACATTCAGGAGTGTGACAAAGATT...AAACAGATGGTGGCTG	1524	
196	pAspHisThrAlaProGlyHisArgThrPheGluThr.....	208	
1525	GAGTCACTGTCTCCATGGTGTCTGCTGTCTGTGACGGTG	1574	
209	.....LysAspGlnProGluTrpAspSer	216	
1575	TGATCACAAAGATCCGTCCTCTGCAACTCCCGCCCGCCAGATGAACGGG	1624	
217	ThrAspGlyGluGlyAsp.....	222	

1625

AAGCCCTGTGAAGTGAAGCCCGGAGACCAAAAGCCTGCAAGAAAGACGC

1674

223

.....TrpSerLeuTrpSerValCysS

230

1675

CTGCCCAATTAAATGGAGGTGGGTCCCTGGTCACCATGGGACATCTGCT

1724

230

erValThrCysGlyAsnGlyAsnGlnLysArgThrArgSerCys.....

244

1725

CTGTCACTCTGGAGGAGGTGCAGAGACGACGCCGACTCTGTAAACAC

1774

245

.....GlyTyrAlaCysThr.....AlaThrGluSe

253

1775

CCACACCCCGAGTTTGAGGCAAGACTGTGTGGCATGTGACAGAAA

1824

253

rArgThrCysAspArgProAsnCysPro

262

1825

TCAAGTTTGAACAAGCAGGACTGCCCA

1852

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA1999.DAT:AAx21356

seq\_documentation\_block:

ID

AAx21356 standard; cDNA; 5412 BP.

XX

AC

AAx21356;

XX

DT

21-MAY-1999 (first entry)

XX

DE

Human BAI2 gene.

XX

XX

Human; BAI2; brain; cancer; drug; diagnosis; prevention; treatment; ss.

XX

XX

Homo sapiens.

XX

XX

Key

Location/Qualifiers

FT

CDS

379..5097

FT

/\*tag= a

/product= "BAI2 protein"

FT

XX

JP11032766-A.

PN

XX

09-FEB-1999.

PD

XX

16-JUN-1997; 97JP-0176485.

PF

XX

23-MAY-1997; 97JP-0150460.

PR

XX

(SAKA ) OTSUKA PHARM CO LTD.

PA

XX

WPI; 1999-183823/16.

DR

XX

P-PSDB; AAW9301.

DR

XX

New human BAI gene - is expressed in brain plays important role in

PT

XX

cancer formation

PT

XX

Claim 4; Page 36-42; 62pp; Japanese.

PS

XX

This sequence represents the gene encoding the human BAI2 protein. The

CC

XX

new gene is expressed specifically in the brain and plays an important

CC

XX

role in cancer formation in the brain. The BAI2 protein can be used in

CC

XX

drug compositions to diagnose, prevent or treat such cancers.

CC

XX

Sequence 5412 BP; 943 A; 1790 C; 1652 G; 1027 T; 0 other;

XX

alignment\_scores:

Quality: 143.00

Length: 457

Ratio: 0.745

Gaps: 19

Percent Similarity: 42.013

Percent Identity: 19.912

alignment\_block:

US-09-863-824-2 x AAx21356

..

Align seg 1/1 to: AAx21356

from: 1 to: 5412

114 SerLysAlaAspIleAsnGlyGln\*\*\*ProAsnIleGlnValThrIleGI 130  
:::|||||:::|||||::: ||| ||| ||| |||  
1120 GCTAGGCCGATTTGCACCTCGGGAGCAGCAATGATGTTCACAACCGA 1169  
  
130 uValValAspGlyProAspSerGluAlaAspLysAspGlnHisProGluA 147  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1170 GATGAGATATTGTGAGGACCCGAA.....GAGAACC CGAAG 1307  
  
147 snLysProSerTrp.....SerValProSerPro..... 156  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1208 TCAAAAACCCAGTGGCCGAGGCTCGAGATGAGCTGGCTATACATGGCG 1257  
  
157 .....AspTrpArgIaIrrPrpGlnArgse 165  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1258 CAGACAGCGACCCCGCGCTGAGAGTGTTGCCCGTGGAGCGTGTGTC 1307  
  
165 rLeuSerLeuAlaArGAlaAsnSerGlyAspGlnAspTy\*\*\*\*\*TyAsp 182  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1308 CCTGACG.....TGTGGCGAGGCTCTCCAGTGGCGACCCGCTCCT 1348  
  
182 erThrSerAspAspSerAsnPheLeu.....Asn 191  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1349 GTGTCTCTCCCCTATGGACCTGTGCAGCGGCCCTCGCGGAGACC 1398  
  
192 ProProArGlyTrpAspHisThrAlaProGlyHisargThrPheGlu... 207  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1399 AGGCCCTGCCAACAAATTACGCCACCTGCCCATGTCACGCCGTGGGAGGA 1448  
  
207 ..... 207  
  
1449 GTGGGGTCTGGAGCCTGTCTCCGACAGTGC GGCGGGGGTCCCGGA 1498  
  
208 .....ThrLysAspGlnProGluTyAspSerThrAspGlyglu 220  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1499 GCCGATGCGACACTGCTGCCCCCCCCAGCACGGCGCAAGCCTGCGAG 1548  
  
221 Gly..... 221  
|||  
1549 GGTCTGAGTGCAGACTAAGCTTCGAGTATGGCTGCCTGCCGGGTGGA 1598  
  
222 .....AspTrpSerLeuTrpSerValCysSerValThrCysG 234  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1599 AGGCCAGTGGTTAGAATGGGTCTGGGGCCCATGCTCCACGTCCTGTG 1648  
  
234 lyAsnGlyAsnGlnLysargThrArgSerCysGlyTyAla..... 247  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1649 CCAATGGGACCCAACAGCGCGACCGCGAAGTTCAGCGTGGCGGCCACGC 1698  
  
248 .....CysThrAla.....ThrCluSerArgThrCysAspArgPr 259  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1699 TGGGCCATGACACGGGTGCCCTCACTGCACACCGGAGTGCACACCT 1748  
  
259 oAsnCysProGlyIleGluAspThrPheArgThrAlaAlaThrLyuVals 276  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1749 CGAGTGGCGGCCACTGATAGCAAGTG..... 1776  
  
276 eLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheGlu... 291  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1777 .....GGGCCATGGAAATGGCTGGACGCTGTGCTTAAG 1809  
  
292 ...ValAspThrAspSerCysGluArgTrpMetSerCysLysSerGluPh 307  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1810 ACGTGTGACACAGGCTGGCAGCGCGCTTCGCGATGTGCCAGGCCACGG 1859  
  
307 eLeuLysLysTyMethIisLysValMetAsnAspLeuProSerCysProc 324  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1860 CAGCAGGGCTAC.....CCCT 1876  
  
324 ysSerTyProThrGluValAlaTySerThrAlaAsp..... 336  
|||:::|||||:::|||||::: ||| ||| ||| |||  
1877 GCGAGGCCACCGGAGGAGGTGAAGCTTGTGATGAGAAGAGGTGTCCA 1926

337 IlePheAspArgIleLysArgLysAspPhe.....ArgTyrLy 349  
||| :::: ||| :::: |  
1927 GCCTTCATAGATGCGAGGCATGACGTGATGCTGATGACGTGA 1976  
349 saspAlaserGlyProLysGliuLysLeuGliulleTyrlYrsProthrAla. 365  
| |||| :::: :::: ::: |  
1977 GAAGGCAGCTGCTGGCGAGATCATCTACAAGAATGCCCCCGAATGCCT 2026  
366 .....ArgTyrCyslleArger..... 371  
2027 CAGGGTGCTCCAGCGCGCGCTCTCTCCTCACTGCCCAAGCGTGGCGCTAC 2076  
371 ..... 371  
2077 TGGGGGCTCCCCAGCTTTTGCTCGTGCTATCTCCCATGACTACCCTACCT 2126  
372 .MetLeuSerLeuGliuSerThrThrLeuAlaAlaglnHisCyscystyrg 388  
||||| :::: :::: ::::  
2127 GTATCTGTCACTTAGGGAGCACCTGGCCAAAGGCGACGCATGCTGGCAG 2176  
388 llyAspAnmNetlInleullethrArGlyLysGliyalGlyThrProasn 404  
llyAsnMettlInleullethrArGlyLysGliyalGlyThrProasn 404  
2177 GCGAGGCATGTCGCGAGTGTGGCG... AGCCTCCAGGAG 2214  
  
405 LeuilleSer.....ThrGlupHeSerAleGliuleHlsTyrlysValas 419  
||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| ::::  
2215 CTACTGGCGCCGCGCACCTACTATAGATGGGACCTGCTCTCTCTGTGGA 2264  
419 pvalleuProTrillelleCylsLysGliyalSrpSerArgTyrAsnGuA 436  
|::::|  
2265 CATTCG.....A 2272  
436 laArgProProAsnAsnGlyGlnLysCystThrGluserProserAspGlu 452  
::: :::: :::: :::: ||||| :::  
2273 GGAATGCTCACTGACACCTTTAAGAGGGCCACCTACGTGCCCTCGCGTGAT 2322  
  
453 AspTyrlleLysGlnPheGln 459  
||| :::: |||||  
2323 GATGTGACGCGCTTCTCCAG 2343

---

OM of: US-09-863-824-2 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Apr 25, 2002 10:16 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09863824/runat\_24042002\_150505\_13882/app\_query.fasta\_1.527  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500  
-FCAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM\_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09863824 -CGN1\_1.89 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-863-824-2  
Query length: 464  
Database: Issued\_Patents\_NA\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 50.240000

score\_list:

Sequence	Strd Orig	zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-862-021B-9	130.00	191.20	0.0058	4029	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-288B-9	130.00	191.20	0.0058	4029	
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US93-03164-9	130.00	191.20	0.0058	4029	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-985-526-4	115.50	178.34	0.0303	1326	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-996-083-2	115.50	164.60	0.1764	4183	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-429-516-2	115.50	164.60	0.1764	4183	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-341-587-4	113.50	157.08	0.4627	5802	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-985-526-2	113.00	182.22	0.0184	657	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2	110.50	72.38	2.4e+04	4403765	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1	110.50	72.36	2.4e+04	4411529	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-808-982-1	107.00	153.19	0.7625	3014	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-306-902A-1	107.00	153.19	0.7625	3014	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-862-021B-13	106.50	158.34	0.3936	1816	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-288B-13	106.50	158.34	0.3936	1816	
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US93-03164-13	106.50	158.34	0.3936	1816	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-862-021B-11	105.50	149.67	1.20	3226	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-288B-11	105.50	149.67	1.20	3226	
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US93-03164-11	105.50	149.67	1.20	3226	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-808-982-3	103.00	146.72	1.75	2831	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-306-902A-3	103.00	146.72	1.75	2831	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1	103.00	58.83	1.3e+05	4411529	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-257-073-14	100.00	147.23	1.64	1725	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-449-645A-14	100.00	140.16	4.05	3116	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-702-367A-14	100.00	140.16	4.05	3116	
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US93-04681-14	100.00	140.16	4.05	3116	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-723-534-3	100.00	135.65	7.23	4547	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-459-101A-1	99.50	151.41	0.9577	1128	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-449-285A-6	99.00	152.44	0.8396	960	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-080-255-4	99.00	134.79	8.07	4201	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-465-713-4	99.00	134.79	8.07	4201	
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US93-05857-4	99.00	134.79	8.07	4201	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-061-376-4	99.00	122.33	39.89	11907	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-320-559-1	99.00	120.18	52.56	14255	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-327-392-1	99.00	120.18	52.56	14255	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-55	99.00	120.18	52.56	14255	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-545-860D-1	99.00	120.18	52.56	14255	
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US94-04486-1	99.00	120.18	52.56	14255	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-206-006-1	98.50	145.42	2.06	1601	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-768-301-3	98.50	139.46	4.43	2635	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-428-517-1	98.50	104.05	415.88	50937	

/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-447-464-2 - 98.00 129.36 16.20  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-716-679-2 - 98.00 129.36 16.20  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-917-722-1 + 97.50 139.30 4.53  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-214-278-6 + 97.50 132.06 11.45  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-068-740A-10 + 97.50 132.06 11.45  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-862-021B-9  
seq\_documentation\_block:  
Sequence 9, Application US/07862021B  
Patent No. 5279966  
GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M  
APPLICANT: Klar, Avihu  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,021B  
FILING DATE: 19920405  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4029 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 226..2647  
US-07-862-021B-9

alignment\_scores:  
Quality: 130.00 Length: 455  
Ratio: 0.722 Gaps: 22  
Percent Similarity: 39.560 Percent Identity: 21.099  
alignment\_block:  
US-09-863-824-2 x US-07-862-021B-9  
Align seg 1/1 to: US-07-862-021B-9 from: 1 to: 4029  
23 ValLeuArgGlySer.....GlyAlaAlaAspGlyProAspAlaAl 36  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1215 ATCTGCAGGAGTCTGTGCACCAAGAGTGTGGCTGGGTCCAGAAAGTGG 1264  
36 aaLaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnAsnLeu.AsnValGly 52  
|||||  
1265 TGCAGGA .....CCTAAATTCCTGGGATGCTGGC 1293  
53 SerAspThrThrSerGluThrSerPheSerLeuSerLysGluAlaProAr 69

```
1294  ACAGGACAC.....GGGTGACCTACAGTCCACAAA 1325
69  gGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgProArgP 86
1326  C.....AAGCCCAAA 1336
86  heArgGlnGluThrClyHisProSerLeuGlnArgAspPheProArgSer 102
1337  TTCTCTCAGGAAAAATCCGACCCCTGACTAGTCTGGACCATCTCCAGAGT 1386
103  PheLeuLeuAspLeuProAsnPheProAsp..... 112
1387  CCTTCTTATGAC.....CCGGAAGTGGTCCATCACACAGT 1424
113  .....LeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
1425  GGCCAGAGTCGTCATCGAGAGAAATGCCGGAAGGGAGAGAAATGCAACA 1474
125  le.....GlnValThrIleGluValValAspGlyProAspSerGlu 138
1475  TTGTACCTGACAATGTGGATGATATGTCGACCTCGCTCCAGAGAG 1524
139  AlaAspLysAspGlnHisProGlu.....AsnLysProSe 150
1525  AAAGATGAAGATGACACCCCTGAACCTGCATCTACTCAACTGGTCCCC 1574
150  rTrp..... 151
1575  ATGGTGGCGCTGCAGCTCTTCCACTTGTGAAAAGGTAAGAGATGCGGC 1624
152  .....SerValProSerProAsp... 157
1625  AACGCATGTGAAGCACAGCTGGACCTCAGTGTCCCTGTCCTGCACACC 1674
157  ..... 157
1675  CAGGACTTCCAGCCCTGCATGGCCCGCGCTGCAGCATGAAGATGGCTC 1724
158  .....TrpArgAlaTrpGlnArgSerLeuSerL 168
1725  CACCTGTACCATGTGCGAGTGCATCACCCTGTCACCTGCAGTGTCTGT 1774
168  euAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSer 184
1775  GTGGCATGGGTATGAGGTCCCGGAGAGGTACGTGAAGCAGTTCCGGAA 1824
185  AspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaPr 201
1825  GACGGCTCGGTGTCATGTGTGCC..... 1848
201  oGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThra 218
1849  .....ACGGAAGACACAGAGAAGTGCAGGTCAACGAGGAGTGT 1888
218  spGlyGluGly.....AspTrpSerLeuTrpSerValCysSer 230
1889  CTCTAGCAGCTGCCTGGTACTAGTGGGGTGGTGGATGACTGACTGCAGC 1938
231  ValThrCysGlyAsnGlyAsnGlnLysArgThrArg..... 242
1939  GCCACCTGTGGATGGCATGAAGAAGCCGACCCGATGATGTCAGATGAG 1988
243  ....SerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrC 256
1989  CCCCGCGGACGGCTCCATGTGCAAGGGGAGACTTCGACGGCGGAGAAAT 2038
256  ysAspArgProAsnCysProGlyIleGluAspThrPheArgThrAlaAla 272
2039  GCATGATGCTGTGATGTCATACCATCCCG..... 2067
273  ThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLe 289
```

```
2068  ....TGCTTGCTGTCTCTTGGTCCGAGTGGAGCGACTGTAGCGT 2108
289  uPheGluValAspThrAspSerCysGluArgTrpMetSerCysLysSerG 306
2109  G.....ACCTGTGGAGGCGCATGCGGACCGCCAGC 2140
306  luPheLeuLys.....Lys 310
2141  GGATGCTCAGTCTCTGGCAGAGCTGGGGAGCTGTAATGAGGATCTGGAG 2190
311  TyrMetHisLysValMetAsnAspLeuProSerCysPro.....CysSe 325
2191  CAGCGCGAGAAGTGTATG.....CTGCCAGATGCCCATTCAGTCGCGA 2234
325  rTyrProThrGluValAlaTyrSerThrAlaAsp..... 336
2235  ACTCAGTGAAGTGTCCAGTGGTCTGTAATGTAACAAGTCTCTGTGGAAAG 2284
337  .....IlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
2285  GTCACATGATTCGAACCCGCAACATCCAAATGGAACCTCAGTTTGGAGGT 2334
351  AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367
2335  GCACCCCTGCCCA...GAGACTGTGCAACGCAAGAGTCCGTCGCCGGAA 2381
367  rCysIleArgSer 371
2382  ATGCTTCGCAGC 2394
seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-313-288B-9
seq documentation block:
; Sequence 9, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; NAME/KEY: CDS
```



LOCATION: 226..2647  
US-08-313-288B-9

alignment\_scores:

Quality: 130.00 Length: 455  
Ratio: 0.722 Gaps: 22  
Percent Similarity: 39.560 Percent Identity: 21.099

alignment\_block:

US-09-863-824-2 x US-08-313-288B-9 ..

Align seg 1/1 to: US-08-313-288B-9 from: 1 to: 4029

```

23 ValLeuArgGlySer.....GlyAlaAlaAspGlyProAspAlaAla 36
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1215 ATCTGCAGAGATCTGTCACCAAGGAGTGGCTGGTGGTCAGAAAGTGG 1264
36 aAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnAsnLeu.AsnValGly 52
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1265 TGCAGGA.....CCTAATTCCCTGGGATGCTGGC 1293
53 SerAspThrThrSerGluThrSerPheSerLeuSerLeuGluAlaProAr 69
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1294 ACGGACAGC.....GGGTGACCTACGAGTACCCAAA 1325
69 gGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgProArgP 86
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1326 C.....AAGCCCA 1336
86 heArgGlnGluThrGlyHisProSerLeuGlnArgAspPheProArgSer 102
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1337 TTCCTCAGGAATAATCCGACCCCTGACTAGTCTGGACCATCTCCAGAGT 1386
103 PheLeuLeuAspLeuProAsnPheProAsp..... 112
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1387 CCTTCTCTATGAC.....CCGGAAGGTGGTCCATCACACAAGT 1424
113 .....LeuSerLysAlaAspIleAsnGlyCln**ProAsnI 125
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1425 GCCCAGAGTCTCATCGAGAGAAATGCCGGAAGGAGAAACAATGCAACA 1474
125 le.....GlnValThrIleGluValValAspGlyProAspSerGlu 138
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1475 TTGTACCTGCACAATGTGATGATATTGTAGCCGACCTGGCTCCAGAGAG 1524
139 AlaAspLysAspGlnHisProGlu.....AsnLysProSe 150
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1525 AAAGATGAAGATGACACCCCTGAAACCTGCATCTACTCCAACTGGTCCCC 1574
150 rTrp..... 151
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1575 ATGTCGGCCTGCAGCTTCCACTTGTGAAAGGTAAGAGGATGGCGC 1624
152 .....SerValProSerProAsp... 157
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1625 AACGCATGCTGAAGGCACAGCTGGACCTCAGTGTCCCTGTCTGCACACC 1674
157 ..... 157
1675 CAGGACTTCCAGCCCTGCATGGGCCCGGCTGCAGCGATGAAGATGGCTC 1724
158 .....TTPArgAlaTrpTrpGlnArgSerLeuSerL 168
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1725 CACCTGTACCATGTCGGAGTGGATCCTGCTGCACCCCTGCAGTGTCTCGT 1774
168 euAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSer 184
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1775 GTGGCATGGGTATCAGGTCCCGGAGGAGTACGTGAAGCAGTTCCTCCGAA 1824
185 AspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaPr 201
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1825 GACGGCTCGTGTGCTGCTGCC..... 1848

```

```

201 oGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrA 218
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1849 .....ACGGAAGACAGAGAGAGTGCACGGTCAACGAGGAGTCT 1888
218 spGlyGluGly.....AspTrpSerLeuTrpSerValCysSer 230
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1889 CTCCTAGCAGCTGCCTGGTACTGAGTGGTGGTGGTGGTGGTGGTGGC 1938
231 ValThrCysGlyAsnGlyAsnGlnLysArgThrArg..... 242
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1939 GCCACCTGTGAATGGGCATGAAGAAGCGCACCCGATGGTCAAGATGAG 1988
243 ....SerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrC 256
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1989 CCCCGCGGACGCTCCATGTGCAAGCGGAGACTTCGACGCGGAGAAAT 2038
256 ysAspArgProAsnCysProGlyIleGluAspThrPheArgThrAlaAla 272
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2039 GCATGATGCCTGAGTGCATACCATCCCG..... 2067
273 ThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLe 289
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2068 .....TGCTTGTCTCTCTCTGTGGTCCGAGTGGACGACTGTAGCGT 2108
289 uPheGluValAspThrAspSerCysGluArgTrpMetSerCysLysSerG 306
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2109 G.....ACCTGTGGGAAGGCGATCGGAGCGGCCAGC 2140
306 luPheLeuLys.....Lys 310
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2141 GGATGCTCAAGTCTCTGCAGAGCTGGGGAGCTGTAATGAGGATCTGGAG 2190
311 TyrMetHisLysValMetAsnAspLeuProSerCysPro.....CysSe 325
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2191 CAGCGGAGAAAGTGTATG.....CTGCCAGAGTCCCGCATTTAGTGGCA 2234
325 rTyProThrGluValAlaTyrSerThrAlaAsp..... 336
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2235 ACTCAGTGCAGTGGTCCCGAGTGTCTGTAATGTAAACAAGTCTGTGGGAAG 2284
337 .....IlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2285 GTCACATGATGTCGAACCGGACAATCCAATGGAACCTCAGTTGGAGGT 2334
351 AlaSerGlyProLysGluLysLeuLysLeuLysProThrAlaArgTy 367
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2335 GCACCCCTGCCCA...GAGACTGTGCACGCAAGAGTCCCGTCCCGGAA 2381
367 rCysIleArgSer 371
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2382 ATGCTTCGCAGC 2394

```

seq\_name: /cgn\_6/ptodata/2/lna/PCTUS\_COMB.seq:PCT-US93-03164-9

seq documentation\_block:

```

: Sequence 9, Application PC/TUS9303164
: GENERAL INFORMATION:
: APPLICANT: Jessell, Thomas M
: APPLICANT: Klar, Avithu
: TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
: TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4029 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 226..2647
; PCT-US93-03164-9

```

```

alignment_scores:
  Quality: 130.00      Length: 455
  Ratio: 0.722        Gaps: 22
  Percent Similarity: 39.560  Percent Identity: 21.099

alignment_block:
US-09-863-824-2 x PCT-US93-03164-9  ..

Align seg 1/1 to: PCT-US93-03164-9 from: 1 to: 4029

23 ValLeuArgGlySer.....GlyAlaAlaAspGlyProAspAlaAl 36
  :: |||||
1215 ATCTCGAGAGGATCTGTCGACCAAGAGGTGGCTGGTCCAGAAAGTGG 1264
  ::|||
36 aAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnAsnLeu..AsnValGly 52
  |||||
1265 TGCAGGA.....CCTAATTCCTGGGATGCTGCG 1293
  ::|||
53 SerAspThrThrSerGluThrSerPheSerLeuSerLysGluAlaProAr 69
  ::|||
1294 ACGGACAGC.....GGGGTGACCTACGATCACCACAA 1325
  ::|||
69 gGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgProArgP 86
  ::|||
1326 C.....AGCCACAA 1336
  ::|||
86 heArgGlnGluThrGlyHisProSerLeuGlnArgAspPheProArgSer 102
  ::|||
1337 TTCTCAGGAAAAATCCGACCCCTGACTAGTGTGGACCATCTCAGAGT 1386
  ::|||
103 PheLeuLeuAspLeuProAsnPheProAsp..... 112
  ::|||
1387 CCTTCTATGAC.....CCGGAAGGTGGTCCATCACACAACT 1424
  ::|||
113 .....LeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
  ::|||
1425 GGCCAGAGTGTCTCGAGAGAAATTCGCCGAAGGAGGAGAAATGCAACA 1474
  ::|||
125 le.....GlnValThrIleGluValValAspGlyProAspSerGlu 138
  ||
1475 TTGTACCTGACAATGTGGATGATATTGTAGCCGACCTGCTCCAGAAAG 1524
  ||
139 AlaAspLysAspGlnHisProGlu.....AsnLysProSe 150

```

```

|||::|||::: |||||
1525 AAAGATGAAGATGACACCCCTGAAACCTGCATCTACTCCAACGTGGTCCCC 1574
  |||
150 rTrp..... 151
  |||
1575 ATGGTCGGCTGCAGCTCTTCCACTTGTGAAAAGGGTAAAGAGGATCGGC 1624
  |||
152 .....SerValProSerProAsp... 157
  |||||
1625 AACGCATGCTGAAGGCACAGCTGAGCTCAGTGTCCCTGTCTCTGACACC 1674
  |||
157 ..... 157
  |||
1675 CAGGACTTCCAGCCCTGCATGGGCCCGGCTGCAGGATGAAGATGGCTC 1724
  |||
158 .....TrpArgAlaTrpTrpGlnArgSerLeuSerL 168
  |||
1725 CACCTGTACCATGTGGGATGGATCACCTGTCACCTGGCAGTGTCTCGT 1774
  |||
168 euAlaArgAlaAsnSerGlyAspGlnAspTyr**TyrAspSerThrSer 184
  ::|||
1775 GTGGCATGGGTATGAGTCCCGGAGAGGTACGTGAAGCAGTTCCCGGAA 1824
  |||
185 AspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaPr 201
  |||
1825 GACGGCTCGGTGTCATGCTGCC..... 1848
  |||
201 oGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrA 218
  |||
1849 .....ACGGAAGACAGACAGAGAAGTGCACGGTCAACGAGGAGTGTCT 1888
  |||
218 spGlyGluGly.....AspTrpSerLeuTrpSerValCysSer 230
  ::|||
1889 CTCCTAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1938
  |||
231 ValThrCysGlyAsnGlyAsnGlnLysArgThrArg..... 242
  |||
1939 GCCACCTGTGAATGGGATCAAGAGCGGCACCGCATGGTCAAGATGAG 1988
  |||
243 ....SerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrC 256
  ::|||
1989 CCCCGCGCAGCGCTCCATGTGTCAAGGCGGAGACTTCGACGCGGAGAAAT 2038
  |||
256 ysAspArgProAsnCysProGlyIleGluAspThrPheArgThrAlaAla 272
  |||
2039 GCATGATGCCCTGATGCCATACCATCCCG..... 2067
  |||
273 ThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLe 289
  |||
2068 .....TGCTTGTCTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2108
  |||
289 uPheGluValAspThrAspSerCysGluArgTrpMetSerCysLysSerG 306
  ::|||
2109 G.....ACCTGTGGAGAGGCGATCGGACGCGCCGAGC 2140
  |||
306 luPheLeuLys.....Lys 310
  ::|||
2141 GGATGCTCAAGTCTCTGCGACAGCTGGGGGACTGTAATGAGATCTGGAG 2190
  |||
311 TyrMetHisLysValMetAsnAspLeuProSerCysPro.....CysSe 325
  ::|||
2191 CAGGGGAGAAAGTGTATG.....CTGCCAGAGTCCCCCATTTGACTGCGA 2234
  |||
325 rTyrProThrGluValAlaTyrSerThrAlaAsp..... 336
  ::|||
2235 ACTAGTGAAGTGGTCCCGAGTGTCTGAATGTAACAAAGTCTGTGGGAAAG 2284
  |||
337 .....IlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
  |||
2285 GTCACATGATTGGAACCCGCAATCCAAATGGAACCTCAGTTTGGAGGT 2334
  |||
351 AlaSerGlyProLysGluLysLeuGluLysTyrLysProThrAlaArgTy 367
  |||

```

2335 GCACCTGCCCCA...GAGACTGTGCAACGCAAGGAAGTGCCTGCCCGGAA 2381

367 rCysIleArgSer 371

||||:|||||

2382 ATGCTTCGCAGC 2394

seq\_name: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:US-08-985-526-4

seq\_documentation\_block:

; Sequence 4, Application US/08985526

; Patent No. 6080728

; GENERAL INFORMATION:

; APPLICANT: Mixson, James A

; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA

; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE

; TITLE OF INVENTION: THERAPY

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Connolly, Bove, Lodge, & Hutz

; STREET: 1220 Market Street, P.O. Box 2207

; CITY: Wilmington

; STATE: Delaware

; COUNTRY: U.S.A.

; ZIP: 19899

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/985,526

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/608,845

; FILING DATE: 16-JUL-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: McMorrow Jr., Robert G

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (302) 658-9141

; TELEFAX: (302) 658-5613

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1326 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-985-526-4

alignment\_scores:

Quality: 115.50 Length: 253

Ratio: 1.121

Percent Similarity: 40.711 Percent Identity: 20.949

alignment\_block:

US-09-863-824-2 x US-08-985-526-4 ..

Align seg 1/1 to: US-08-985-526-4 from: 1 to: 1326

193 ProArgGlyTyrAspHisThrAlaProGlyHisArgThrPheGluThrLy 209

||||| |||

208 CTGCTGTGTTG..... 219

209 sasplnProGluTyrAspSerThrAspGlyGluClyAspTrpSerLeut 226

||||| |||

220 .....CCGACGACTCTGGGAGGATGGCTGCTCCATGGTCCGAGT 262

226 rpSerValCysSerValThrCysGlyAsnGlnLysArgThrArg 242

||||| |||

263 GGACCTCTGTCTACGAGCTGTGGCAATGGAATTCACGCGCGCGC 312

243 SerCys.....GlyTyrAlaCys.....ThrAlaThrGluSerAr 254

||||| |||

313 TCCTGCGATAGCCTCAACAACCGATGTGAGGCTCCTCGGTCCAGACAG 362

||||| |||

254 gThrCysAspArgProAsnCysProGlyIleGluAspThrPheArg.... 269

||||| |||

363 GACCTGCCACATTCAGGAGTGT.....GACAAAAGATTAAACAGG 403

270 .....ThrAlaAlaThrGluValSerLeuLeuAla 279

||||| |||

404 ATGCTGGCTGGAGCCACCTGGTCCCGTGTCTATCTTGTCTGTGACATGT 453

280 GlySerGluGluPheAsnAlaThrLysLeuPhe.....G1 291

||||| |||

454 GGTGATGTGTGATCACAAAGATCGCGTCTGCAACTCTCCAGCCGCA 503

291 uValAspThrAspSerCysGlu.....ArgTrpMetSerCysL 304

||||| |||

504 GATGAATGGGAACCTGTGAAGCGGAGGAGGAGGAGGAGGAGGAGGAG 553

304 ySerGluPheLeuLysLysTyrMetHisLysValMetAsnAspLeuPro 320

||||| |||

554 AGAAGAC..... 561

321 SerCysProCysSerTyrProThrGluValAlaTyrSerThrAlaAspI 337

||||| |||

562 GCCTGCCCATCAAT..... 576

337 ePheAspArgIleLysArgLysAspPheArgTrpLysAspAlaSerGlyP 354

||||| |||

577 .....GGAGGCTGGGTC 589

354 roLysGluLysLeuGluIleTyrLysProThr..... 364

||||| |||

590 CTGTGCACCATGGGACATCTTCTGTCCCTGTGGAGGAGGAGGAGGAG 639

365 .....AlaArgTyrCysIleArgSerMetLeuSerLeuGluSerThrTh 379

||||| |||

640 AAACGTAGTCTCTGTGCTGCGACTCTAGAATGACTGAAGAGACAAAGA 689

379 rLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThra 396

||||| |||

690 GTTGCC..... 696

396 rgGlyLys.GlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAl 412

||||| |||

697 ..AATGAGCTGAGCGGCTCCCTATCTATCATCAACGAGTTCAGTAC 744

412 aGluLeu 414

||||| |||

745 AGAAATA 751

seq\_name: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:US-08-996-083-2

seq\_documentation\_block:

; Sequence 2, Application US/08996083A

; Patent No. 6124095

; GENERAL INFORMATION:

; APPLICANT: Magna, Holly

; APPLICANT: Schaffer, Paul

; APPLICANT: Lawton, Michael

; APPLICANT: Yocum, Sue

; APPLICANT: Mitchell, Peter

; APPLICANT: Hutchinson, Nancy

; APPLICANT: Murry, Lynn E.

; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHODIOLASE-2

; FILE REFERENCE: PF-0420 US

; CURRENT APPLICATION NUMBER: US/08/996,083A

; CURRENT FILING DATE: 1997-12-22

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 4183

; TYPE: DNA

seq name: /cgn2 6/otodata/2/ina/6B COMB. seq: US-09-429-516-2

```

190 euAsnProProArgGly...Tip.....AspHisThrAlaProGly... 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
395 TGAACCCACCGCGGCTTCGTGGTCCCAACCGGAGCAACCGCGTGGC 444
203 .....HisArgThrPheGluThrLysAspGlnProG1 213
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 CGCGCGTCTCCAACTACCACTGCGCTTCGCTGCCCACTAGAAAGCTC 494
213 uTyAspSerThrAspGlyGluGlyAspTrpSerLeuTrpSerValCysS 230
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 GTGG.....GGCGGTGGGCGCGGTGGGCTCCCTGCT 526
230 erValThrCysGlyAsnGlnLysArgThrArgSerCys..... 244
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
527 CGGGAGCTGTGGCCAGCGCTGCTGCGCGCGCCCACTGCCCAAGC 576
245 .....GlyTyrAlaCysThrAla.....ThrGluSerArgThrCysAs 257
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
577 CCCGCTGGGATGGGTGTCCCGGCGTCTCTGGAGGCGCAGAAAGTGGT 626
257 pArgProAsnCysProGly 263
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
627 CGGCGCTCGGTGTCCAGGG 645

```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-341-587-4

```

seq_documentation_block:
; Sequence 4, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; EARLIER FILING DATE: 1999-08-31
; EARLIER FILING DATE: PCT/DE98/00096
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 5802
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-341-587-4

```

```

alignment_scores:
Quality: 113.50 Length: 230
Ratio: 1.102 Gaps: 15
Percent Similarity: 44.783 Percent Identity: 26.522

```

alignment\_block:  
US-09-863-824-2 x US-09-341-587-4 ..

Align seg 1/1 to: US-09-341-587-4 from: 1 to: 5802

```

143 GlnHisProGluAsn.....LysProSerTrpSe 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1463 CAGCACAGTGAAGACGCTGGTGTCTCTGCTCAGTCCCACTCTGCTG 1512
152 rValProSerProAspTrp.....ArgA 160
|:||||| ||||| ||||| ||||| ||||| ||||| |||||
1513 GACGCCAGTCCAGACACATGCGCACCATCACCTTGGCTGCATCGACAG 1562
160 laTrpTrpGlnArgSerLeuLeuAlaArgAlaAsnSerGlyAspGln 176
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1563 TAGGATCTGAATCCAGTTTGGCCCTGAGGCTGGTGAATGAGGTGACAGG 1612
176 ..... 176
1613 TGTGAGGCGCGAGTGGAGGTCTTATACCAAGGCTCTTGGGCGACCGTGTG 1662

```

```

177 .AspTyr**TyrAspSerThrSerAspSerAsnPheLeu.....A 191
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1663 CGATCAGCTGGGAC.....ACCAATGATGCCAATGCTCTCTGCAGGC 1706
191 snProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGlu 207
::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1707 AACCGGGCTGTGGTGGGCCATGTACGCCCGCAGGAAATGCCCGTTGGT 1756
208 ThrLysAspGlnPro...GluTyAspSerThrAspGlyGluGlyAspTr 223
::: ||| ||||| ||||| ||||| ||||| ||||| |||||
1757 CAGGCTCAGGACCACTTCTCTGGATGATGTGGCTCTCAGGACACGA 1806
223 pSerLeu...TrpSerVal.....CysSerValThrCysG 234
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1807 GTCTTACCGTGGAGCTGCCCCACCAATGGCTCTCTCCACCAACTGTG 1856
234 LysnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1857 GCCAT.....AGTGAAGACGCTGTGTCTCATCTGTCTCAGT 1891
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluAspTh 267
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1892 TCCAGTCCCGGCCAACACCTAGTCCA.....GACAC 1923
267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluP 284
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1924 TTGGCCAACTCAGATGCATCAACA.....GCAGATCTGAATCCA 1964
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCysGlu...Arg 299
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1965 GTTTGGCCCTGAGCTGGTGAAT...GGAGGTGACAGGTGTGAGGCGCGA 2011
300 .TrpMetSerCysLysSerGluPheLeuLysLysTyrMethHisLysValM 316
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2012 GTGGAGGTCC.....T 2022
316 eAsnAspLeuProSerCysProCysSerTyrProThr 328
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
2023 ATACCGAGGCTCTGGGCGCACCGTGTGTGATGACTACT 2060

```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-985-526-2

```

seq_documentation_block:
; Sequence 2, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-985-526-2

alignment_scores:
  Quality: 113.00      Length: 74
  Ratio: 2.825        Gaps: 3
  Percent Similarity: 54.054  Percent Identity: 33.784

alignment_block:
US-09-863-824-2 x US-08-985-526-2 ..
Align seg 1/1 to: US-08-985-526-2 from: 1 to: 657
193 ProArgGlyThrPaspHisThrAlaProGlyHisArgThrPheGluThrLy 209
||||| |||
208 CCTCGCTGTGG..... 219
209 saspGlnProGluTyAspSerThrAspGlyGluGlyAspTrpSerLeut 226
||||| : : : : : ||||| : : : : : |||||
220 .....CCACGGACTCGCGGAGATGGTGGTCTCCATGGTCCGAGT 262
226 rpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysArgThrArg 242
||||| : : : : : ||||| : : : : : |||||
263 GGACCTCCTGTTCTACGAGCTGTGCAATGGAATTCAGCAGCGCGCGGC 312
243 SerCys.....GlyTyraLaCys.....ThrAlaThrGluSerAr 254
||||| : : : : : ||||| : : : : : |||||
313 TCTGTGGATAGCTCAACACGATGTGAGGCTCTCGTCCAGACAGC 362
254 gThrCysAspArgProAsnCys 261
||||| : : : : : |||||
363 GACCTGCCACATTCAGGAGTGT 384

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-08-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
;
US-09-103-840A-2

alignment_scores:
  Quality: 110.50      Length: 421
  Ratio: 0.639        Gaps: 23
  Percent Similarity: 41.093  Percent Identity: 22.090
```

```

alignment_block:
US-09-863-824-2 x US-09-103-840A-2 ..
Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765
31 AspGlyProAspAlaAlaGly.....AsnAlaSerGlnAlaG1 44
: : : : : ||||| : : : : : |||||
30705 GAAGGCGCCGACGAGGACGCGGGCTGGGCAAGTCAGAATCGCCACGCA 30754
44 nLeuGlnAsnAsnLeuAsnValGlySerAspThrThrSerGluThrSerP 61
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
30755 GGCTCCGACGACCGTGACGCTG.....CCCAACGGTGACGCGTGA 30795
61 heSerLeuSerLysGluAlaProArgGluHisLeuAspHisGlnAlaA1 77
: : : : : ||||| : : : : : ||||| : : : : : |||||
30796 CCGGGCCAGTCCCGAGCTCGCGCGCGATCAAGGCGGCGGCACGCGGC 30845
77 aHisGlnProPheProArgPheArgGlnGlnGluThrGlyHisProS 94
||||| : : : : : ||||| : : : : : |||||
30846 ACACCGATCGAGATGGTTCACACACAGGAA.....TTGCCAT 30886
94 erLeuGlnArgAspPheProArgSerPheLeuLeuAspLeuProAsnPhe 110
||||| : : : : : ||||| : : : : : |||||
30887 CCCGCTACCGGAA.....CCGCGGTGC.....CCAACCCGTCGACT 30924
111 ProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnIleGlnVa 127
||||| : : : : : ||||| : : : : : |||||
30925 CCG..... 30927
127 lThrIleGluValValAspGlyProAspSerGluAlaAspLysAspGlnH 144
||||| : : : : : ||||| : : : : : |||||
30928 .....CCCGATCTCAGCGGGAGACGAGTAGTGT 30955
144 isProGluAsnLysPro....SerTrp.....SerValProSer 155
: : : : : ||||| : : : : : |||||
30956 GTTCACCGCCAGCCCTTGCCCTTGGCCCTAGCAAGCTCTTCTGGACGG 31005
156 ProAspTrpArgAlaTrpTrpGlnArgSerLeuSerLeuAlaArgAla... 171
||||| : : : : : ||||| : : : : : |||||
31006 CCAGAT.....TCAACACATCTCAGCGGTGGGAGGCCAA 31040
172 .....AsnSerGlyAspGlnAspTyr***TyrAsps 182
: : : : : ||||| : : : : : ||||| : : : : : |||||
31041 ACTTTCAGGCTGGATATATCCAGCGCGCGCGACCGCGCGCGCGAGG 31090
182 erThrSerAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHis 198
: : : : : ||||| : : : : : |||||
31091 ACCGAAGCACCGACACCAAC..... 31110
199 ThrAlaProGlyHisArgThrPheGluThrLysAspGlnProGluTyraS 215
||||| : : : : : ||||| : : : : : |||||
31111 ...CAGGCGCGCGCGCTCGATAGGTACT...GACCGCGCGGTC..... 31149
215 pSerThrAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerVal. 231
: : : : : ||||| : : : : : ||||| : : : : : |||||
31150 .ACAACAGAGGAGACACGGGATGACAGATCGAATTCAGTCGACGCTGC 31198
232 .....ThrCys 233
||||| : : : : : |||||
31199 ACATTTAGTTCAGGCGCTGCCCATCACCAGACAGACCGCGCGACTACCTGC 31248
234 GlyAsnGlyAsnGlnLysArgThrArgSerCys..... 244
: : : : : ||||| : : : : : ||||| : : : : : |||||
31249 GGACCGTCCGTCGTCGCGACGCGGATCCGGAAAGTCTGGACTCGCTG 31298
245 .....GlyTyraLaCysThrAlaT 251
||||| : : : : : ||||| : : : : : |||||
31299 GGGCTATTTTTCAGTCAGTCCGCGACACCGCGCGCTGAGTCTCGAGCT 31348
251 hrGluSer..... 253
: : : : : |||||
31349 CAGAAAGCAGTCTCTACAGCAAGCGCGCAACACCGCGCGATATTGCC 31398
```

```

254 ArgThrCysAsp...ArgProAsnCysProGlyIleGluAspThrPheAr 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31399 AGAAGCTCGAAGCTGCGCCGCGATGTGGAGACGACGAGCGAGCGCG 31448
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 gThrAlaAlaThrGluValSerLeuAlaGlySerGluGlu.....P 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31449 TCGCGACGCTCGGCAACATCATTTGACGGGAGCCGATGACAGGCGGATGA 31498
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
284 hAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrp 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31499 CCGACGCCAATCCCGCTTTGACACGGTCCACCCAGCGGCGCACATTCTT 31548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 Met...SerCysLysSerGluPheLeuLysLysTyrMetHisLysValme 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31549 GTTCGGTCTCGCGCGGT.....GGATACATGCATACGCGTCTC 31586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316 tAsnAspLeuProSerCysProCysSerTyrProThr.....GluVala 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31587 GCTGAGCGAGGCGGCGATGGAGACCGACGACGAAACCTCGCGGAAGCCA 31636
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 laTyrSerThrAlaAspIlePheAspArgIleLysArgLysAspPheArg 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31637 TCCTGCTCACC CGCCGACGTG..... 31656
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 TrpLysAspAlaSerGlyProLysGluLysLeuGluIleTyrLysProTh 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31656 ..... 31656

364 rAlaArgTyrCysIleArgSerMetLeuSerLeuGluSerThrThrLeuA 381
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31657 .....TCCTGCCTTAAAGCTTCTGGAAGTACGCAACGAGATCGTGG 31699
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 laAlaGlnHis 384
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31700 CGCGCGGCCAC 31710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-103-840A-1
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIA
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

alignment_scores:
  Quality: 110.50      Length: 421
  Ratio: 0.639         Gaps: 23
  Percent Similarity: 41.093      Percent Identity: 22.090

alignment_block:
US-09-863-824-2 x US-09-103-840A-1 ..

Align seg 1/1 to: US-09-103-840A-1 from: 1 to: 4411529

31 AspGIVpVroAspAlaAlaGlv.....AsnAlaSerGlnAlaG1 44

```

30724 GAAGCGCGCAGCAGCAGCGCGGCTGGCAAGTCAGATCCGCACCCGA 30773  
 44 lLeuGlnAsnAsnLeuAsnValGlySerAspThrThrSerGluThrSerP 61  
 30774 GGCTCCGACGACCGGTGACGCTG.....CCCAACGCTGAGACGGTGA 30814  
 61 heSerLeuSerLysGlu..AlaProArgGluHisLeuAspHisGlnAlaAl 77  
 30815 CCGCGCGCAGTCCCGAGCTCGCGCGGGGATCAAGCGCGCGCCAGCGCG 30864  
 77 ahisGlnProPheProArgProArgPheArgGlnGluThrGlyHisProS 94  
 30865 ACACCGATCGCAGATCGGTTCCAAACACAGGGA.....TTGCCAT 30905  
 94 erLeuGlnArgAspPheProArgSerPheLeuLeuAspLeuProAsnPhe 110  
 30906 CCGCTACCGGAA...CGCGGTG.....CCAACCCCGTCGACC 30943  
 111 ProAspLeuSerLysAlaAspIleAsnGlyGln\*\*ProAsnIleGlnVa 127  
 30944 CG.....CG..... 30946  
 127 lThrIleGluValValAspGlyProAspSerGluAlaAspLysAspGlnH 144  
 30947 .....CCCGGATCTCAGCGGAGACAGTAGGTGT 30974  
 144 lsProGluAsnLysPro...SerTrp.....SerValProSer 155  
 30975 GTTACCGCCAGCGCCTTGCCTTGCCTAGCAAGCTCTCTCGACGG 31024  
 156 ProAspTrpArgAlaTrpTrpGlnArgSerLeuSerLeuAlaArgAla... 171  
 31025 CCAGAT.....TCAACACATCTCAGCGCTGAGGAGGCCAA 31059  
 172 .....AsnSerGlyAspGlnAspTyr\*\*\*TyrAspS 182  
 31060 ACTTTTAGGCTGGATACATCCAGCGGGGACCGCAGCGCGCGCGAGG 31109  
 182 erThrSerAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHis 198  
 31110 ACCGAAGCACCAGCACCAAC..... 31129  
 199 ThrAlaProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAs 215  
 31130 ...CAGGCGGCGCGCTCGATAGTACT...GACCGCCCGTC..... 31168  
 215 pSerThrAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerVal.. 231  
 31169 .ACAACAGAGAGACAGCGGATGACATCGAATTCACGTGACGCCCTGC 31217  
 232 .....ThrCys 233  
 31218 ACATTTAGCTCAGCGCGTGCCCATCACAGCAGACCGCCGATACCTGC 31267  
 234 GlyAsnGlyAsnGlnLysArgThrArgSerCys..... 244  
 31268 GGACGTCGCTGTCGACAGCGGATCCGCGAAGTCTGGACTCGCTG 31317  
 245 .....GlyTyrAlaCysThrAlaAl 251  
 31318 GGGCCTATTTTCAGTGAGTCCGCGACACCGGGCGTGAGCTGCTGAGCT 31367  
 251 hrGluSer..... 253  
 31368 CAGAAGACGATGCTACACAGCAGCAAGCCGACAAACACCGCCGATATGCC 31417  
 254 ArgThrCysAsp...ArgProAsnCysProGlyIleGluAspThrPheAr 269  
 31418 AGAACCTGGAAACGTGCGCGCGGATGTGGGACGACACAGGACGCGCGG 31467  
 269 gThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGlu.....P 284





398 sGlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuH 415  
: : : : :  
1324 GATGACCCAGCCCAAGTCCAGTCTCTAATGTCACCTGCTCAGGCC 1373  
415 iSTyRLysValAspValLeuProTrpIleIleCysLysGlyAspTrpSer 431  
: : : : :  
1374 ACTGGGAGTGGCGCCATA..... 1393

432 ArgTyrAsnGluAlaArgProPro 439  
: : : : :  
1394 CGTTGCACACAGCTCACCACCT 1417.

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-306-902A-1

seq\_documentation\_block:  
; Sequence 1, Application US/09306902A  
; Patent No. 6277585  
; GENERAL INFORMATION:  
; APPLICANT: Tessier-Lavigne, Marc  
; Leonardo, E. David  
; Hink, Lindsay  
; Masu, Masayuki  
; Kazuko, Keino-Masu  
; TITLE OF INVENTION: Netrin Receptors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/306,902A  
; FILING DATE: 07-May-1999  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UC96-217  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: nucleic acid  
; LENGTH: 3014 base pairs  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-306-902A-1

alignment\_scores:  
Quality: 107.00 Length: 292  
Ratio: 1.039 Gaps: 15  
Percent Similarity: 35.274 Percent Identity: 21.575

alignment\_block:  
US-09-863-824-2 x US-09-306-902A-1 ..  
Align seg 1/1 to: US-09-306-902A-1 from: 1 to: 3014

219 GlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysGlyAs 235  
: : : : :  
730 GGGTGGTGGAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 779

235 nGlyAsnGlnLysArgThrArgSer..... 243  
: : : : :  
780 TGGCTGGCAGAAACGGAGCGAGCTGCACCAACCCGCGACCTCTCAACG 829  
243 ..... 243  
830 GGGCGGCTTCTGTGAGGGGAGAAATGTCCAGAAACAGCCTGCGCCACT 879  
244 .....Cy 244  
880 CTGTGCCCCAGTGGATGGAGCTGGAGTTCGTGGAGTAAGTGGTCAGCCTG 929  
244 sGlyTyrAlaCysThrAlaThrGluSerArgThrCysAspArgPro.... 259  
: : : : :  
930 TGGGCTTGACTGCACCCACTGGCGAGCGCGAGTGTCTTGACCCAGCAC 979  
260 .....AsnCysProGlyIleGluAspThrPheArgThr 270  
: : : : :  
980 CCCGCAATGGAGGTGAGGAGTGTGGGGTGTGACCTGGACACCCGCAAC 1029  
271 AlaAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaTh 287  
: : : : :  
1030 TGTACCACTGACCTCTGCCTG..... 1050  
287 rLysLeuPheGluValAspThrAspSerCys...Glu.ArgTrpMetSer 302  
: : : : :  
1051 .....CACACCGCTTCTTGGCCCGAGGACGTGGCTCTCT 1084  
303 .....CysLysSerGluPheLeuLysLysTy 311  
: : : : :  
1085 ACATCGGCTTGTGCTGTGGCTGTGTGCCTCT..... 1117  
311 rMethHisLysValMetAsnAspLeuProSerCysProCysSerTyrProT 328  
1118 .....TCTTGC...TGTTCGTGGCCC. 1135  
328 hrGluValAlaTyrSerThrAlaAspIlePheAspArgIleLysArgLys 344  
: : : : :  
1136 .....TTGGACTCATTTACTGTGCGCAAGAGGAAG 1165  
345 AspPhe.....ArgTrpLys..... 349  
: : : : :  
1166 GGCTGGACTCGGATGTGGCGGACTGTCTCATCTCACCTCGGGCTTCCAG 1215  
350 .....AspAlaSerGlyProLysGluLysLeuGluIleTyrLysProThrA 365  
: : : : :  
1216 CTGTGCACATCAAGCCCGCA.....AAGCAGACAA 1247  
365 laArgTyrCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAla 381  
: : : : :  
1248 CCCCCACCTGCTCACCACCTCAGC...CAGACCTCAGCACCACCTACCA 1294  
382 AlaGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLy 398  
: : : : :  
1295 CCTACCGAGCGAGTCTAT.....GTTCCGAGGCG 1323  
398 sGlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuH 415  
: : : : :  
1324 GATGACCCAGCCCAAGTTCAGCTCTCTAATGTCACCTGCTCAGGCC 1373  
415 iSTyRLysValAspValLeuProTrpIleIleCysLysGlyAspTrpSer 431  
: : : : :  
1374 ACTGGGAGTGGCGCCATA..... 1393  
432 ArgTyrAsnGluAlaArgProPro 439  
: : : : :  
1394 CGTTGCACACAGCTCACCACCT 1417

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-862-021B-13  
seq\_documentation\_block:  
; Sequence 13, Application US/07862021B  
; Patent No. 5279966

```

; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1816 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1705
; US-07-862-021B-13

```

```

alignment_scores:
  Quality: 106.50      Length: 329
  Ratio: 0.813        Gaps: 14
  Percent Similarity: 39.818  Percent Identity: 21.277

```

```

alignment_block:
  US-09-863-824-2 x US-07-862-021B-13

```

```

Align seg 1/1 to: US-07-862-021B-13 from: 1 to: 1816
93 ProSerLeuGlnArgAspPheProArgSerPheLeu...LeuAspLeuPr 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 CCCACCATCCCGAGGATAAATCCGACCTCTGACAAAGTCTGGATCACC 435
108 oAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 ACAAGCCCTTCTATGACCAGAGGT.....GGGCCAATCATACATATAG 479
125 leGlnValThrIleGlu..... 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 CTCGAGTGTGGATTGAAGGATTGCCAGGAGGAGACAGTGCATATATT 529
131 .....ValValAspGlyProAspSerGluAl 139
530 ATACCCGACAACTGGATGACATAGTAGCAGATCTGGTAACGGAAGAGAA 579
139 aAspLysAspGlnHisProGlu.....AsnLysProSerT 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 AGACGAGATGATACCCCGAGACCTGCATATATTCCAACTGGTCCCT 629

```

```

151 rp..... 151
||
630 GGTGGGCTGCAGCTCGGCCACTCGCACAGGCAAGCGGATGAGACAG 679
152 .....SerValProSerProAsp..... 157
|||||:|||||:|||||:|||||:|||||:|||||:
680 CGCATGTTAAAGGCTCAGTTAGATCTCAGTGTCTCCTGCCAGACACTCA 729
157 ..... 157
730 GGACTTTGAACCCTGCATGGGCGCGCTGCAGCGATGACGAAGCCTCTA 779
158 .....TTPArgAlaTTPTrpGlnArgSerLeuSerLeu 168
|||||:|||||:|||||:|||||:|||||:|||||:
780 CTTGCATGATGTCAGAAATGATCCTGTCGCGCTGCAGCGCTCCTGC 829
169 AlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSerAs 185
|||||:|||||:|||||:|||||:|||||:|||||:
830 GGGATGGGAATTGAGTTCAGGAGAGATACGTCACGAGTCCCAAGA 879
185 pAspSerAsnPheLeuAsnProProArgGlyTTPAspHisThrAlaProG 202
|||||:|||||:|||||:|||||:|||||:|||||:
880 CGGTTCCTTGTGTAAGTCCCA..... 901
202 lyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrAsp 218
|||||:|||||:|||||:|||||:|||||:|||||:
902 .....ACGGAAGAACTGAGAAATGCAATTCATGAGGAGTGTGAG 943
219 GlyGluGly.....AspTTPSerLeuTTPSerValCysSerVa 231
|||||:|||||:|||||:|||||:|||||:|||||:
944 CCAAGACGCTGTATGTCACGGAATGGCAGAGTGGGAGGAGTGCAGCGC 993
231 lThrCysGlyAsnGlnLysArg.....ThrA 242
|||||:|||||:|||||:|||||:|||||:|||||:
994 TACATCGCGATGGTATGATGAAGCGGCACAGGATGATAAGATGACTC 1043
242 rgSerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrCys 256
|||||:|||||:|||||:|||||:|||||:|||||:
1044 CAGCGGATGGATCTATGTGCAAGCCGACACACAGAGGTGAGAATGC 1093
257 AspArgProAsnCysProGlyIleGluAspThrPheArg.....ThrAl 271
|||||:|||||:|||||:|||||:|||||:|||||:
1094 ATGATGCCCGATGTCATACCATCCCGTGTGTCCTGCTGCTGA 1143
271 aAlaThrGluValSerLeuAlaGlySerGluGluPheAsnAlaThrL 288
|||||:|||||:|||||:|||||:|||||:|||||:
1144 ATGGAGTGATTGTCAGCGCTTACCTGTGCAAGCCGACACAGACAGAGA 1193
288 ysLeuPheGluValAspThrAspSerCysGluArgTTPMetSerCysLys 304
|||||:|||||:|||||:|||||:|||||:|||||:
1194 GAATGTTGAAGTCCCGTCTGAACTTGGAGAT.....TGCAAT 1231
305 SerGluPheLeuLysLysTyrMetHisLysValMetAsnAspLeuProSe 321
|||||:|||||:|||||:|||||:|||||:|||||:
1232 GAGGAAGTGGAACTGAAACAAAGTGGAAAGTGCATG.....CTTCCTGA 1275
321 rCysProCysSerTyrPro.....ThrGluValAlaTyr 332
|||||:|||||:|||||:|||||:|||||:|||||:
1276 ATGCCCTATAAGCTGTGAATTGACAGAGTGTGCTTAC 1312

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-288B-13
seq_documentation_block:
; Sequence 13, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and AviHu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas

```

```

; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1816 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1705
; US-08-313-288B-13

alignment_scores:
  Quality: 106.50      Length: 329
  Ratio: 0.813         Gaps: 14
  Percent Similarity: 39.818   Percent Identity: 21.277

alignment_block:
US-09-863-824-2 x US-08-313-288B-13  ..

Align seg 1/1 to: US-08-313-288B-13 from: 1 to: 1816

93  ProSerLeuGlnArgAspPheProArgSerPheLeu...LeuAspLeuPr 108
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386  CCCACCATCCCGAGGATAAATCCGACCTCTGACAGTCTGGGATCACC 435

108  oAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
436  ACAAAGCCCTTCTATGACCAGGT.....GGCCCAATCATACCTATAG 479

125  leGlnValThrIleGlu..... 130
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
480  CTCGAGTGTGATTGAAGGATTGCCAGGAGGAGAACAGTGCATATT 529

131  .....ValValAspGlyProaspSerGluAl 139
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
530  ATACCCGACAACTGGATGACATAGTACAGATCTGGTAACGGAGAGAA 579

139  aAspLysAspGlnHisProGlu.....AsnLysProSerT 151
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
580  AGACAAGATGATACCCCGGAGACCTGCATATATCCAACTGGTCCCT 629

151  rp..... 151
    ||
630  GGTGGCTGTCAGCTCGCCACCTGCGACAAGGCGGAGGATGAGACAG 679

152  .....SerValProSerProasp..... 157
    |||:|||||:|||||:|||||:|||||:|||||:
680  CGCATGTTAAAGGCTCAGTTAGATCTCAGTGTCCCTGCCCGACACTCA 729

157  ..... 157

```

```

730  GGACTTTTGAACCCCTGCATGGGCGCGCTGCAGCGATGACGAGCCCTCTA 779
158  .....TrpArgAlaTrpTrpGlnArgSerLeuSerLeu 168
    |||:|||||:|||||:|||||:|||||:|||||:
780  CCTGCATGATGTCAGAAATGGATACCTACCTGGTCGCGTCGACGCGCTCCTGC 829
169  AlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSerAs 185
    |||:|||||:|||||:|||||:|||||:|||||:
830  GGGATGGGAATTGAGTTCAGGAGAGATACGTCAAGCAGTTCCCGAGAGA 879
185  pAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaProG 202
    |||:|||||:|||||:|||||:|||||:|||||:
880  CGGTTCTCTGTGTAAGTCCCA..... 901
202  lyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrAsp 218
    |||:|||||:|||||:|||||:|||||:|||||:
902  .....ACGGAAGAACTGAGAAATGCATTTGTCAATGAGGAGTGTGAG 943
219  GlyGluGly.....AspTrpSerLeuTrpSerValCysSerVa 231
    |||:|||||:|||||:|||||:|||||:|||||:
944  CCAAGCAGCTGTATAGTCACGGAATGGCAGAGTGGGAGGAGTGCACGCG 993
231  lThrCysGlyAsnGlyAsnGlnLysArg.....ThrA 242
    |||:|||||:|||||:|||||:|||||:|||||:
994  TACATGCGGATGGTATGAAGAAGCGCACAGGATGATAAAGATGACTC 1043
242  rgSerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrCys 256
    |||:|||||:|||||:|||||:|||||:|||||:
1044  CAGCGGATGGATCTATGTGCAAAAGCCGACACACAGAGGTTGAGAAATGC 1093
257  AspArgProAsnCysProGlyIleGluAspThrPheArg.....ThrAl 271
    |||:|||||:|||||:|||||:|||||:|||||:
1094  ATGATGCCGAATGTATACCATCCCGTGGTGTTCCTTGGTCTCTGA 1143
271  aAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrL 288
    |||:|||||:|||||:|||||:|||||:|||||:
1144  ATGGAGTGTGTCAGCGTTACCTGTGCAAAAGCCAGCACCAACAGACAGA 1193
288  ysLeuPheGluValAspThrAspSerCysGluArgTrpMetSerCysLys 304
    |||:|||||:|||||:|||||:|||||:|||||:
1194  GAATGTTGAAGTCCCGCTCTGAACTTGGAGAT.....TGCAAT 1231
305  SerGluPheLeuLysLysTyrMetHisLysValMetAsnAspLeuProSe 321
    |||:|||||:|||||:|||||:|||||:|||||:
1232  GAGAACTGGAACTGAAACAAGTGGAAAGTGCATG.....CTTCTCTGA 1275
321  rCysProCysSerTyrPro...ThrGluValAlaTyr 332
    |||:|||||:|||||:|||||:|||||:|||||:
1276  ATGCCCTATAAGCTGTGAATTTGACAGAGTGTGCTTAC 1312

seq_name: /cqn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03164-13

seq_documentation_block:
; Sequence 13, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1816 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1705
; PCT-US93-03164-13

alignment_scores:
Quality: 106.50 Length: 329
Ratio: 0.813 Gaps: 14
Percent Similarity: 39.818 Percent Identity: 21.277

alignment_block:
US-09-863-824-2 x PCT-US93-03164-13 ..
Align seg 1/1 to: PCT-US93-03164-13 from: 1 to: 1816

93 ProSerLeuGlnArgAspPheProArgSerPheLeu...LeuAspLeuPr 108
|||||:|||||:|||||:|||||:|||||:
386 CCACCATTCCTCCAGGATAAATCCGCTCTGACAAAGTCGTGGATCACC 435
108 oAnPheProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
|||||:|||||:|||||:|||||:|||||:
436 ACAAGCCCTCTATGACCAGAGGT.....GGGCCAATCATACCTATAG 479

125 leGlnValThrIleGlu..... 130
|||||:|||||:|||||:|||||:
480 CTCAGTTGTGATTGAAGGATTGCCAGGAGGAGACAGTGCATATT 529

131 .....ValValAspGlyProAspSerGluAl 139
|||||:|||||:|||||:|||||:
530 ATACCCGACACGCTGGATGACATAGTAGCAGATCTGGTAACGGAAGAA 579

139 aspLysAspGlnHisProGlu.....AsnLysProSerT 151
|||||:|||||:|||||:|||||:
580 AGACGAAGATGATACCCCGAGACCTGCATATATTCCTGCTCCCTCC 629

151 rp..... 151
630 GGTCGGCTCGACCTCGGCCACCTCGGCACAGGGGCAAGCGGATGACAG 679

152 .....SerValProSerProAsp..... 157
|||||:|||||:|||||:|||||:
680 CGCATGTTAAAGGCTCAGTTAGATCTCAGTGTTCCTCCCTGCCAGAC 729

157 ..... 157
730 GGACTTTGAACCTCGCATGGGCGCGCTGCACGCGATGACGAAGCCTCTA 779

158 .....TTPArgAlaTriPTrpGlnArgSerLeuSerLeu 168
|||||:|||||:|||||:|||||:
780 CCTGCATGATGTCAGATGGATCACCTGGTCCGCGTGCAGCGCTCCTGC 829

```

OM of: US-09-863-824-2 to: EST:\* out\_format : pfs

Date: Apr 25, 2002 9:43 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2_1/USP70_spool/US09863824/runat_24042002_150505_13849/app_query.fasta_1.527  
-DB=EST -QFWT=fastap -SURFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -HOCAL=LOCAL -OUTFMT=pls -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09863824@cgn1_1.4938  
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-863-824-2

Query length: 464

Database: EST:\*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1532.850000

score\_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
gb_est2:BF472502	-	1103.00	1790.72	1.7e-90	644	1 BE472502 UI-M-BH3-axd-h-03-0-UI
gb_est2:BG162342	+	1077.00	1747.41	4.4e-88	691	1 BG162342 de3e08.yl Wellcome CR
gb_est1:AW873508	+	1034.00	1711.67	4.3e-86	583	1 AW873508 ho0h09.xl Soares_NFL
gb_est2:BF606213	+	968.00	1569.90	3.4e-78	645	1 BJ060213 BJ060213 NIBB Mochil
gb_est2:BF302850	+	854.00	1378.97	1.5e-67	968	1 BF302850 602032745F1 NCI_CGAP
gb_est2:BB947736	+	827.50	1341.99	1.7e-65	544	1 BE947736 UI-M-BH3-axd-h-03-0-UI
gb_est2:BT891452	+	703.50	1136.96	4.4e-54	666	1 BT891452 2F637-3-000707 Zebra
gb_est1:AA884248	+	684.50	1110.84	1.3e-52	425	1 AA884248 am32b09.sl Soares_NFL
gb_est2:BT170419	+	678.50	1095.28	9.2e-52	715	1 BT170419 zfs1TARAMA00029 Zebra
gb_est1:AW426875	+	671.00	1089.91	1.8e-51	383	1 AW426875 61687 MARC 3BOV Bos ta
gb_est2:BM191035	+	619.00	1003.18	1.2e-46	446	1 BM191035 daj31b08.yl NICHD XGC
gb_est2:WB6257	+	594.00	964.78	1.7e-44	356	1 WB6257 zdg91f09.r1 Soares fetal
gb_est1:AW455535	+	574.50	932.20	1.1e-42	379	1 AW455535 UI-M-BH3-avb-e-02-0-UI
gb_est2:BT535390	+	564.00	910.93	1.7e-41	549	1 BT535390 398882 MARC 4BOV Bos U
gb_est1:AL545675	+	525.00	840.18	1.5e-37	1032	1 AL545675 AL545675 LTI_NFL006.F
gb_est1:AW435642	+	496.00	801.23	2.2e-35	479	1 AW435642 74577 MARC 1PIG Sus sc
gb_est2:BF571346	+	491.00	789.15	1.0e-34	682	1 BF571346 602077485F1 NIH_MGC.62
gb_gss:CN5021K6	+	485.50	776.92	5.0e-34	914	1 AL199023 Tetraodon nigroviridis
gb_est2:BG037854	+	478.50	772.98	8.3e-34	463	1 BG037854 dc69a05.yl NICHD XGC H
gb_est1:BB322233	+	472.00	758.72	5.2e-33	643	1 BB322233 BB322233 RIKEN full-le
gb_est1:AL659826	+	467.00	749.94	1.6e-32	679	1 AL659826 AL659826 XGC-neurula S
gb_est1:AW435639	+	449.00	730.85	1.8e-31	266	1 AW435639 74571 MARC 1PIG Sus sc
gb_est2:W14682	+	354.50	575.93	7.8e-23	275	1 W14682 mb30h04.r1 Soares mouse
gb_est2:BJ071721	+	342.00	546.72	3.3e-21	608	1 BJ071721 BJ071721 NIBB Mochil
gb_est2:BG739780	+	337.00	535.15	1.5e-20	827	1 BG739780 602630502F1 NCI_CGAP
gb_est1:AA447294	+	327.50	520.74	9.3e-20	747	1 AA447294 AC.E04R PH001-Normal H
gb_est2:BE287313	+	308.00	489.67	5.0e-18	693	1 BE287313 601093984F1 NCI_CGAP.M
gb_est2:BT1442837	+	277.00	441.60	2.4e-15	546	1 BT1442837 dag19a08.yl NICHD XGC
gb_est1:AW435643	+	256.00	422.20	2.9e-14	141	1 AW435643 74579 MARC 1PIG Sus sc
gb_gss:A0606601	+	255.50	404.13	2.9e-13	673	1 A0606601 HS.2095.B1.E07 MR CIT
gb_est1:BB586477	+	242.50	391.39	1.5e-12	311	1 BB586477 BB586477 RIKEN full-le
gb_est1:AW435648	+	213.00	352.52	2.2e-10	133	1 AW435648 74587 MARC 1PIG Sus sc
gb_est2:BU082998	+	211.00	330.88	3.5e-09	702	1 BU082998 BJ082998 NIBB Mochil
gb_est2:BT1449830	+	206.00	328.04	5.0e-09	433	1 BT1449830 dag32a03.y3 Blackshear
gb_gss:AZ995663	+	168.50	263.31	2.0e-05	589	1 AZ995663 2M0281N06R Mouse 10kb
gb_est1:AT1430783	-	150.00	240.58	0.0004	298	1 AT1430783 hb66a04.xl Soares mouse
gb_est1:AT1070670	-	140.00	222.35	0.0039	353	1 AT1070670 UI-R-C2-mx-c-05-0-UI.S
gb_est1:AA464532	+	131.50	203.45	0.0437	555	1 AA464532 zx85a05.sl Soares ovari
gb_est2:HZ4875	-	129.00	212.30	0.0141	172	1 HZ4875 Y142h10.r1 Soares breast
gb_est1:AA243836	-	129.00	198.78	0.0796	585	1 AA243836 zr67g12.r1 Soares_NHNM

gb\_gss:AG041438 - 128.50 196.94 0.1008 642 1 AG041438 Pan troglodytes DN  
gb\_est2:BF769031 - 128.00 199.20 0.0755 486 1 BF769031 ILO-IT0033-101100-1  
gb\_est1:AA243572 + 127.50 195.91 0.1151 608 1 AA243572 zr67g12.sl Soares  
gb\_est1:AI278172 + 127.00 198.66 0.0809 440 1 AI278172 ql77h07.xl Soares  
gb\_est1:AI866968 + 127.00 197.95 0.0885 469 1 AI866968 wnl3f10.xl NCI\_CGAI

seq\_name: gb\_est2:BF472502

seq documentation block:

LOCUS BF472502 644 bp mRNA linear EST 04-DEC-2000

DEFINITION UI-M-BH3-axd-h-03-0-UI.r1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone

UI-M-BH3-axd-h-03-0-UI 5', mRNA sequence.

ACCESSION BF472502

VERSION BF472502.1 GI:11541685

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 644)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 97044477

MEDLINE

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone Distribution:

Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It

should be noted that Bento Soares is generating a small number of

additional specialized non-redundant arrays of BMAP cDNAs whose

availability will be considered under appropriate and limited

collaborative arrangements

Seq primer: M13 Reverse

Location/Qualifiers

1..644

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-BH3-axd-h-03-0-UI"

/clone\_lib="NIH\_BMAP\_M\_S4"

/env\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; The

NIH\_BMAP\_M\_S4 library is a subtracted library of a series,

ultimately derived from a mixture of individually tagged

normalized libraries from ten regions of the mouse brain

(cerebellum, brain stems, olfactory bulbs, hypothalamus,

cortex, amygdala, basal ganglia, pineal gland, striatum,

hippocampus) after a series of subtractions to reduce the

representation of cDNAs from which ESTs had already been

generated. The following serially subtracted libraries

were generated in this process: NIH\_BMAP\_M\_S4,

NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,

NIH\_BMAP\_M\_S3. The subtracted library

(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified

cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and

NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived

was used as a driver in a hybridization with a pool of

the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1

libraries in the form of single-stranded circles. The

remaining single-stranded circles (subtracted library)

was purified by hydroxyapatite column chromatography,

converted to double-stranded circles and electroporated

into DH10B bacteria (lifetechologies) to generate the

NIH\_BMAP\_M\_S4 library. This procedure has been previously

described (Bonaldo, Lennon and Soares, Genome Research





```

seq_name: gb_est2:BJ060213
seq_documentation_block:
LOCUS      BJ060213      645 bp      mRNA      linear      EST 11-DEC-2001
DEFINITION laevis cDNA clone XL065c05 5', mRNA sequence.
VERSION    BJ060213.1 GI:17496943
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 645)
AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
TITLE      Expressed genes in X. laevis embryo
JOURNAL    Unpublished (2001)
COMMENT    Contact: Tadasu Shin-i
            Center for Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshin@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
            1..645
             /organism="Xenopus laevis"
             /db_xref="taxon:8355"
             /clone="XL065c05"
             /clone_lib="NIBB Mochii normalized Xenopus tailbud
             library"
             /tissue_type="whole embryo"
             /dev_stage="stage 25"
BASE COUNT 194 a 147 c 170 g 133 t 1 others
ORIGIN

alignment_scores:
  Quality: 968.00      Length: 214
  Ratio: 4.990         Gaps: 1
  Percent Similarity: 90.654      Percent Identity: 79.907

alignment_block:
US-09-863-824-2 x BJ060213      ..
Align seg 1/1 to: BJ060213 from: 1 to: 645
136 AspSerGluAlaAspLysAspGlnHisProGluAsnLysProSerTrpSe 152
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
4 GACTCGGCGCGGAAATCAATGCAAAAGGAGAACATGCCAGCTGGCC 53
152 rValProSerProAspTrpArgAlaTrpTrpGlnArgSerLeuSerLeuA 169
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
54 GGTTCATCACCTGACTGAGAGAGCTGGTGGCAGAGGTGACCACTTGC 103
169 laArgAlaAsnSerGlyAspGlnAspTyr**TyrAspSerThrSerAsp 185
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 CCGCTATGAATTCAGGGGACAGGATTACAATACGACAGCACTACTAG 153
186 AspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaProGl 202
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 GACAGCAACTTCTTAACCTCTGGGAGGAGGAATAGACAGTGGCCAG 203
202 yHisArgThrPheCluThrLysAspGlnProGluTyrAspSerThrAspG 219
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 CCACCGAACATTTGATACAAAGGACAGCGCTGAGTATGATGATG 253
219 lyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysGlyAsn 235
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 GTGAGGAGATTCGACAGTTCGTCAGTGCAGTGCACCTGCTGTGATGC 303

```

---

```

236 GlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAlaThrGl 252
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 GGTAAACAAAAGCGCACCGAACCCTGTGCTATGCTGTACAGCGACAGA 353
252 uSerArgThrCysAspArgProAsnCysProGlyIleGluAspThrPheA 269
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 GTCAGAAGAACTTGTGATATGCCAAACTGTCCGGAATCGAAGACACATTC 403
269 rgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsn 285
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 GAACAGCAGCCACTGAAGTCAGCTTCTGCGAGGAATGAGGATTTCAT 453
286 AlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrpMetSe 302
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 GCTACAAAAGCTGTTTGTGTTGATCTACACAGCTGTGACGCGATGGATGA 503
302 rCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAsnAspL 319
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
504 CTGCAAAAGTGAATTCCTCAAAAAGTATATGCACAAAGTGGCTAATGACC 553
319 euProSerCysProCysSerTyrProThr.GluValAlaTyrSerThrAl 335
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
554 TTCCAGCTGCCCTTGTCTCTACCCAACTTGAAGTGGCTTATAGCACCGC 603
335 aAspIlePheAspArgIleLysArgLysAspPheArgTrp 348
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
604 TGAATATACGATCGGATCAACGCAAAAGACTTTTCGATGG 643

```

```

seq_name: gb_est2:BF302850
seq_documentation_block:
LOCUS      BF302850      968 bp      mRNA      linear      EST 21-NOV-2000
DEFINITION 602032745F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4167658 5',
            mRNA sequence.
ACCESSION  BF302850
VERSION    BF302850.1 GI:11249408
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 968)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            CDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9458 row: j column: 11
            High quality sequence stop: 726.
FEATURES   Location/Qualifiers
            1..968
             /organism="Mus musculus"
             /strain="FVB/N"
             /db_xref="taxon:10090"
             /clone="IMAGE:4167658"
             /clone_lib="NCI_CGAP_SG2"
             /lab_host="DHI0B (T1 phage-resistant)"
             /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
             NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
             dt. Average insert size 1.3 kb. Constructed by Life
             Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 298 a 242 c 219 g 209 t
ORIGIN

alignment_scores:

```







```

seq_name: gb_est1:AA884248

seq_documentation_block:
LOCUS      AA884248                425 bp      mRNA      linear      EST 04-JAN-1999
DEFINITION  am32b09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1468505 3', mRNA sequence.
ACCESSION  AA884248
VERSION    AA884248
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 425)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapb-r@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert length: 1007 Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 424.

FEATURES             source
     source          1..425
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1468505"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NBHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-727811, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."

BASE COUNT  113 a 140 c 98 g 72 t 2 others
ORIGIN

alignment_scores:
Quality: 684.50      Length: 141
Ratio: 5.225         Gaps: 1
Percent Similarity: 92.908 Percent Identity: 92.908

alignment_block:
US-09-863-824-2 x AA884248 ..
Align seg 1/1 to: AA884248 from: 1 to: 425

64 SerLysGluAlaProArgGluHisLeuAspHisGlnAlaAlaHisGlnPr 80
|||||
3 TCCAAGAGACCAAGGAGGATCTGGACCACCAAGGCTGCACACCAACC 52
|||||

80 oPheProArgProArgPheArgGlnGluThrGlyHisProSerLeuGlnA 97
|||||
53 CTTCGCCAGACCCGATCCGACAGAGACGGGGCACCT.TCATTGCAAA 101
|||||

97 rgAspPheProArgSerPheLeuLeuAspLeuProAsnPheProAspLeu 113
|||||
102 GAGATTCCCCAGATCTTTCTCTGTGATCTACCAAACTTTCCAGATCTT 151
|||||

114 SerLysAlaAspIleAsnGlyGln***ProAsnIleGlnValThrIleGl 130
|||||
152 TCCAAGGTGATPATCAATGGGAGAAATCCAAATATCCAGGTCACCATAGA 201
|||||

```

```

130 uValValAspGlyProAspSerGluAlaAspLysAspGlnHisProGluA 147
|||||
202 GGTGTCGACGGTCTGACTCTGAAGCAGATAAAGATCAGATCCGGAGA 251
|||||

147 snLysProSerTrpSerValProSerProAspTrpArgAlaTrpTrpGln 163
|||||
252 ATAAGCCAGCTGGTCTCAGTCCCATCCCGACTGGCGGCCTGGTGCCAG 301
|||||

164 ArgSerLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAspTy***Ty 180
|||||
302 AGGTCCCTGTCTTGGCAGGCAACACAGCGGGACCAAGACTACAAGTA 351
|||||

180 rAspSerThrSerAspAspSerAsnPheLeuAsnProProArg...GlyT 196
|||||
352 CGACAGTACCTCAGACGACGCAACTTCTCAACCCCGCCCGCTGGNG 401
|||||

196 rpAspHisThrAlaProGlyHis 203
|||||
402 GACCCATACAGCCCGCAGGCCAC 424

seq_name: gb_est2:BI704199

seq_documentation_block:
LOCUS      BI704199                715 bp      mRNA      linear      EST 19-SEP-2001
DEFINITION  zfshtARAMA000029 zebrafish shield stage whole embryo cDNA library
MPMGp637 Danio rerio cDNA clone MPMGP637H1826 5', mRNA sequence.
ACCESSION  BI704199
VERSION    BI704199.1 GI:15680329
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE  1 (bases 1 to 715)
AUTHORS   Dickmeis,T., Aanstad,P., Clark,M., Fischer,N., Herwig,R., Mourrain
            P., Blader,P., Rosa,F., Lehrach,H. and Strahle,U.
TITLE     Identification of Nodal signaling targets by array analysis of
            induced complex probes
JOURNAL   Dev. Dyn. (2001) In press
COMMENT   Contact: Hennig S
            Laboratory 123, Dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            selected as regulated by Taram-A* signalling by cDNA array analysis
            High quality sequence stop: 715.

FEATURES             Location/Qualifiers
     source          1..715
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="MPMGp637H1826"
                     /clone_lib="zebrafish shield stage whole embryo cDNA
                     library MPMGP637"
                     /tissue_type="whole embryo"
                     /dev_stage="shield stage, 6 hrs post-fertilisation"
                     /lab_host="E.coli, XLI blue MRP"
                     /note="vector: pSport1; Site_1: NotI; Site_2: SalI;
                     oligo-dT-NotI primed, SalI adaptors, directionally cloned,
                     library normalised by oligonucleotide fingerprinting"

BASE COUNT  194 a 171 c 207 g 142 t 1 others
ORIGIN

alignment_scores:
Quality: 678.50      Length: 223
Ratio: 3.833         Gaps: 3
Percent Similarity: 79.372 Percent Identity: 62.332

alignment_block:

```

```

US-09-863-824-2 x BI704199 ..
Align seg 1/1 to: BI704199 from: 1 to: 715

94 SerLeuGlnArgAspPheProArgSerPheLeuLeuAspLeuProAsnPh 110
|||||
48 AGCGTCGAGGAGACATCCCTGGAGCTTCTGCTGGACCTGCAGAACTT 97
|||||
110 eProAspLeuSerLysAlaAspIleAsnGlyGln**ProAsnIleGlnV 127
|||||
98 TCCTGATCTCTCCAAAGCAGATATAATGGCGAAGACCCCAATATACAGG 147
|||||
127 alThrIleGluValValAspGlyProAsp...SerGluAlaAspLysAsp 142
|||||
148 TCACCATAGAGTGTGGACAGCTGGAGAGTTCTGAGCCAGAGAGGGA 197
|||||
143 GlnHisProGluAsnLysProSerTrpSerValProSerProAspTrpAr 159
|||||
198 ATGCGTAAGAAACAAAGCCTGGCTGGCGAGCTCTCTAGC.....TGGAG 241
|||||
159 galatrpTrpGlnArgSerLeuSerLeuAlaArgAlaAsnSerGly.... 174
|||||
242 GAACCTGGTGGCAGCTTCATCTCCTCTCATCTCTCTCTCTCTCTCTCT 291
|||||
175 .....AspGlnAspTyr**TyrAspSerThrSerAspAsp 186
|||||
292 CGAAGGGGCGCTCAGGAACAGGATTACCCCTACGAGAGCAACACAGAGAC 341
|||||
187 SerAsnPheLeuAsnProArgGlyTrpAspHisThrAlaProGlyHi 203
|||||
342 AGCAACTTCTCTCAAGCCACTCGGAGACTGGGAGAGAGTGAAGAGTGA 391
|||||
203 sargThrPheGluThrLysAspGlnProGluTyrAspSerThrAspGlyG 220
|||||
392 AGCTGAGCGGAGAGCAAGCAACCCAGACTGAATACACTACATAGACGGG 441
|||||
220 luGlyAspTrpSerLeuTrpSerValThrCysGlyAsnGly 236
|||||
442 AAGCGGACTGGAGTCTTGGTCTCCATGCAGTGTATCTCTGTGGAATGG. 490
|||||
237 AsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAlaThrGluSe 253
|||||
491 AACCAGAGCGGACCAAGATCGTGGCGTTATGCTGCACCTGCACACAGAGTC 540
|||||
253 rArgThrCysAspArgProAsnCysProGlyIleGluAspThrPheArgT 270
|||||
541 ACCGACATGTGACATCCACAGCTGTCAGTGATTGAAGATGATTCAAGA 590
|||||
270 hrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAla 286
|||||
591 CGGCAGCAACTGAAGTCAGTTGCTGCCTAGCACCTGTGAGTCAATGCA 640
|||||
287 ThrLysLeuPheGluValAspThrAspSerCysGluArgTrpMetSerCy 303
|||||
641 NCAGAGCTGTTTGGAGTGTATACCGACAGCTGGGACGCTTGATGACTGC 690
|||||
303 LysSerGluPheLeuLys 309
|||||
691 AGAGTGAGTTTTATAGGAAA 709
|||||

seq_name: gb_est1:AW426875

seq_documentation_block:
LOCUS AW426875 383 bp mRNA linear EST 25-APR-2001
DEFINITION 61687 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW426875
VERSION AW426875.1 GI:6954822
KEYWORDS EST.
SOURCE
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovinae; Bos.

```

1 (bases 1 to 383)  
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett  
, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,  
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and  
Keele, J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 20  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCACAGTCACGAGC  
Plate: 26 row: L column: 24  
Seq primer: ATTTAGGTGACACTATAG.  
FEATURES  
    source  
        1..383  
            /organism="Bos taurus"  
            /db\_xref="taxon:9913"  
            /clone\_lib="MARC 3BOV"  
            /tissue\_type="pooled"  
            /lab\_host="DH10B"  
            /note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;  
            Library made from pooled tissue from marrow, alveolar  
            macrophage, ovary, fetal semitendinosus muscle, and fetal  
            longissimus muscle."  
BASE COUNT 97 a 104 c 113 g 69 t  
ORIGIN

alignment\_scores:  
    Quality: 671.00 Length: 123  
    Ratio: 5.455 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.374

alignment\_block:  
US-09-863-824-2 x AW426875 ..  
Align seg 1/1 to: AW426875 from: 1 to: 383

233 CysGlyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysTh 249  
|||||  
14 TGTGGGAATGGCAACCAAGAAACGACGAGGTCCTTGTGGCTACGCGTGTAC 63  
|||||  
249 rAlaThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluA 266  
|||||  
64 TGCAACCGAATCTAGGACGCTGTGACCGCTCCAGCTGCCAGGAATTGAAG 113  
|||||  
266 spThrPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGlu 282  
|||||  
114 ACACCTTCAGGACAGCTGCCACTGAAGTGAAGTCTGCTTGGCGGAAGTGA 163  
|||||  
283 GluPheAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluAr 299  
|||||  
164 GAGTTTAAATGCCAACAACTGTTTGAAGTCGACACGAGCTGTGAGCG 213  
|||||  
299 gTrpMetSerCysLysSerGluPheLeuLysLysTyrMetHisLysValM 316  
|||||  
214 CTGGATGAGTTGCAAGAGCGAGTTCTCTAAAGAAATACATGATAGGTGA 263  
|||||  
316 etAsnAspLeuProSerCysProCysSerTyrProThrGluValAlaTyr 332  
|||||  
264 TCAACGACCTGCTAGCTAGCTGCCCTCTCTCTACCCACCAGGAGTGCCTAC 313

127 alThrIleGluValAlaAspGlyProAspSerGluAlaAspLysAspGln 143

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library was constructed by M.Patima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung Nhlh19k."

BASE COUNT 88 a 109 c 92 g 61 t 6 others  
ORIGIN

alignment\_scores:  
Quality: 594.00 Length: 118  
Ratio: 5.351 Gaps: 0  
Percent Similarity: 94.068 Percent Identity: 93.220

alignment\_block:

US-09-863-824-2 x W86257 ..

Align seg 1/1 to: W86257 from: 1 to: 356

284 PheAsnAlaThrLysLeuPheGluValAspThrSerCysGluArgTr 300  
|||||  
2 TTTAATGCCACCAAACTGTTTGAGTTGACACAGACGCTGTGAGCNC 51

300 pMetSerCysLysSerGluPheLeuLysLysTyrMethLysValMetA 317  
|||||  
52 GATGAGCTCAAAAGCAGTTCTTAAAGAGTACATGCACAAAGTGATGA 101

317 snAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSer 333  
|||||  
102 ATGACCTGCCAGCTGCCCTCTCTACCCACCTGAGTGCCCTACAGC 151

334 ThrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrPlysAs 350  
|||||  
152 ACGGCCNACATCTTCGACCTCATCAAGCGCAAGGACTTCGCTGGAAGGA 201

350 pAlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgT 367  
|||||  
202 CGCCAGCTGGCCCAAGGAGAACTGGAGATCTACAAGCCCACTGCCCGT 251

367 yrcysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGln 383  
|||||  
252 ACTGATCGCTCCATCGCTGCTCCCTGGAGAGCACCACGCTGCGNCACAG 301

384 HisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAl 400  
|||||  
302 CACTGCTGTACGGCGACAACTGCACATGCACCTCATCACCAGGNCAGGNGTC 351

400 agly 401

|||

352 GGCG 355

seq\_name: gb\_est1:AW495535

seq\_documentation\_block:

LOCUS AW495535 379 bp mRNA linear EST 24-FEB-2000  
DEFINITION UI-M-BH3-avb-e-02-0-UI-s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
UI-M-BH3-avb-e-02-0-UI 3', mRNA sequence.

ACCESSION AW495535

VERSION AW495535.1 GI:7065816

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 379)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 9704477

# COMMENT

Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov  
Oligo-qt track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=No.

# FEATURES

source

Location/Qualifiers

1..379

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-BH3-avb-e-02-0-UI"

/clone\_lib="NIH\_BMAP\_M\_S4"

/dev\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73b-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)

TAG\_LIB=NIH\_BMAP\_M\_S4

TAG\_TISSUE=hippocampus

TAG\_SEQ=TTGAC"

BASE COUNT 99 a 122 c 89 g 69 t

# ORIGIN

alignment\_scores:

Quality: 574.50

Ratio: 4.910

Percent Similarity: 91.406

Percent Identity: 83.594

alignment\_block:

US-09-863-824-2 x AW495535 ..

Align seg 1/1 to: AW495535 from: 1 to: 379

36 AlaAlaGlyAspAlaSerGlnAlaGlnLeuGlnAsnAsnLeuAsnValG1 52

|||||

3 GCCCGCCGACGTCAGCGGGTCCAGTTCGAGTAATACCTCAACCTGGA 52

|||||

52 ySerAspThrThrSerGluThrSerPheSerLeuSerLysGluAlaProA 69

```

BASE COUNT      140 a      144 c      138 g      127 t
ORIGIN
embryos."
alignment_scores:
  Quality:      564.00      Length:      112
  Ratio:        5.127      Gaps:      1
  Percent Similarity: 98.214      Percent Identity: 93.750
alignment_block:
  US-09-863-824-2 x BI535390      ..
  Align seg 1/1 to: BI535390 from: 1 to: 549
293 AspThrAspSerCysGluArgTrpMetSerCysLysSerGluPheLeu.L 309
   ::::::::::::::::::::
214 AACACGGACAGCTGTGATCGCTTGTGATGAGTTGCAGACGCGAGTTCTCTATA 263
   ::::::::::::::::::::
309 ySLysTyrrMethiSLysValMetAsnAspLeuProSerCysProCysSer 325
   ::::::::::::::::::::
264 AGAATAATACATGATTAAGGTGATCAACAGCACTGCCTAGCTGCCCTGCTGCC 313
   ::::::::::::::::::::
326 TyrProThrGluValAlaTyrSerThrAlaAspIlePheAspArgIleLys 342
   ::::::::::::::::::::
314 TACCCACCGAGTGGCCCTACAGCACAGCTGACATCTTCGACCGCATCAA 363
   ::::::::::::::::::::

```

364 GCGCAAGGACTTCGCTGGAGGAGCCAGCGGGCCCAATGAGAAGCTGG 413  
359 luiletTyrLysProThrAlaArgTyrCysIleAArgSerMetLeuSerLeu 375

```

376 GluSerThrThrlieuAlaIaGlnHisCysCysTyrGlyAspAsnMetGl 392
|||||
464 GAGAGACACACGCTGGCTGCCCATCACTGTTGCTACGGCGACACATGCA 513
|||||
392 nleullethrArgGlyLysGlyAlaGlyThrPro 403
|||||
514 GCTTATCACAGGGGCAAAAGGGGGGGAGGCC 547
|||||

seq_name: gb_est1.AL545675
seq_documentation block:

```

<b>sequence_documentation_block:</b>					
LOCUS	AL545675	1032 bp	mRNA	linear	EST 16-FEB-2001
DEFINITION	AL545675 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI009YJ17 5 prime, mRNA sequence.				
ACCESSION	AL545675				
VERSION	AL545675.1	GI:12878157			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1032)				
AUTHORS	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.				
TITLE	Full-length cDNA libraries and normalization				

Genoscope - Centre National de Séquençage  
BP 191 91006 EVRY cedex - France  
Email: [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr) Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

```
FEATURES
Location/Qualifiers
source
    l..1032
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_xref="CS01009yJ7"
    /clone_lib="LTI_NFL006_PL2"
    /tissue_type="placenta"
    /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2002, 14:46:33 ; Search time 72.05 Seconds  
(without alignments)  
1114.083 Million cell updates/sec

Title: US-09-863-824-2  
Perfect score: 2517  
Sequence: 1 MVRLAAELLGLLGLLTLH.....CTSPSDEDIKQFEAREY 464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2160	85.8	391	4 Q9H599	Q9h599 homo sapien
2	926	36.8	658	4 Q95432	Q95432 homo sapien
3	321	12.8	60	4 Q9BQL4	Q9bql4 homo sapien
4	137.5	5.5	1121	5 Q966P9	Q966p9 caenorhabdi
5	137.5	5.5	1148	5 Q966Q0	Q966q0 caenorhabdi
6	135	5.4	1637	6 Q9XSV8	Q9xsv8 bos taurus
7	133.5	5.3	461	5 Q9S522	Q9s522 drosophila
8	133.5	5.3	685	6 Q9RTS5	Q9rts5 bos taurus
9	133.5	5.3	763	5 Q9XZD0	Q9xzd0 drosophila
10	131.5	5.2	660	5 Q23832	Q23832 cryptospori
11	127	5.0	765	5 Q9U8J9	Q9u8j9 neospora ca
12	125.5	5.0	873	5 Q9V692	Q9v692 drosophila
13	124	4.9	2165	5 Q19791	Q19791 caenorhabdi
14	122.5	4.9	1280	11 Q9EPX2	Q9epx2 mus musculu
15	120.5	4.8	4123	4 Q75851	Q75851 homo sapien
16	120	4.8	822	4 Q9UGT4	Q9ugt4 homo sapien

function

ALIGNMENTS

RESULT 1

Q9H599 ID Q9H599 PRELIMINARY; PRT; 391 AA.

AC Q9H599; 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE BA149118.1 (NOVEL PROTEIN) (FRAGMENT).

GN BA149118.1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Wilton S.;

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL133463; CAC16127.2; -

DR InterPro; IPR000884; TSP1.

DR Pfam; PF00090; tsp\_1; 1.

DR SMART; SM00209; TSP1; 1.

DR PROSITE; PS50092; TSP1; 1.

FT NON\_TER 1

SQ SEQUENCE 391 AA; 44556 MW; 3B44913173C95649 CRC64;

Query Match 85.8%; Score 2160; DB 4; Length 391;

Best Local Similarity 99.5%; Pred. No. 9.8e-201;

Matches 399; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 74 HQAAHQFPFRFQETGHPSLQDRFPRSELLDLPNPPDLKADINQXQNIQVTIEWD 133

Db 1 HQAAHQFPFRFQETGHPSLQDRFPRSELLDLPNPPDLKADINQXQNIQVTIEWD 60

QY 134 GP0SEADKQHPKPSWSVPSPDRAWQWQSLARANSQDQDYXYDSTSDSSNFLNPP 193

Db 61 GP0SEADKQHPKPSWSVPSPDRAWQWQSLARANSQDQDYKYDYDSTSDSSNFLNPP 120

QY 194 RGMWDHTAPGHTTETKDKQPEYDSTDGDGWSLWSVCSVTGNGNQRKTRSCGYACTATES 253

NO function

[illegible]

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission."  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006615; AAK68231.1;  
SQ SEQUENCE 1121 AA; 126159 MW; 151B03F47E5BF1BC CRC64;

Query Match 5.5%; Score 137.5; DB 5; Length 1121;  
Best Local Similarity 24.4%; Pred. No. 0.0004;  
Matches 47; Conservative 26; Mismatches 71; Indels 49; Gaps 10;

QY 130 EVVDGPDSEADKQ-----HPENKPSWSVSPDWRAMQWSLSLARANSQDQDYXYS--- 182  
DB 471 ECIDGVSCGASEKILCNQPCPEWS-----QWTAW-----TVCDERCGEESIRLRNKC 521  
QY 183 -TSDSNFLNPPRGWDHTAPGHRFTETKQDPEYDSTDGEGDWSLWSVCTCGNGNQRT 241  
DB 522 MNAENNNACDGP-----AQDQMSCPYRDCPKWE-----EWGEWADCSITCGGQTKRL 569  
QY 242 RSC--GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVDVDFSC 297  
DB 570 RKCDSGNECSGPGEMRFQIASCPYWGDTWPSGCSVSCGQ-----VC 614  
QY 298 ERWMSK-KSEFLK 309  
DB 615 ERTRKICITDEFLQ 627

RESULT 5  
Q966Q0 PRELIMINARY; PRT; 1148 AA.  
AC Q966Q0; DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL PROTEIN C3687.5B.  
GN C3687.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Minx P., Minx M.;  
RT "The sequence of C. elegans cosmid C3687.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission."  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006615; AAK68232.1;  
SQ SEQUENCE 1148 AA; 129319 MW; 3D0714E2A42D202D CRC64;

Query Match 5.5%; Score 137.5; DB 5; Length 1148;  
Best Local Similarity 24.4%; Pred. No. 0.00042;  
Matches 47; Conservative 26; Mismatches 71; Indels 49; Gaps 10;

QY 130 EVVDGPDSEADKQ-----HPENKPSWSVSPDWRAMQWSLSLARANSQDQDYXYS--- 182  
DB 471 ECIDGVSCGASEKILCNQPCPEWS-----QWTAW-----TVCDERCGEESIRLRNKC 521  
QY 183 -TSDSNFLNPPRGWDHTAPGHRFTETKQDPEYDSTDGEGDWSLWSVCTCGNGNQRT 241  
DB 522 MNAENNNACDGP-----AQDQMSCPYRDCPKWE-----EWGEWADCSITCGGQTKRL 569  
QY 242 RSC--GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVDVDFSC 297  
DB 570 RKCDSGNECSGPGEMRFQIASCPYWGDTWPSGCSVSCGQ-----VC 614  
QY 298 ERWMSK-KSEFLK 309  
DB 615 ERTRKICITDEFLQ 627

RESULT 6  
Q9XSV8 PRELIMINARY; PRT; 1637 AA.  
AC Q9XSV8; DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE SCO-SPONDIN (FRAGMENT).  
GN SCO-SPONDIN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gobron S., Creveaux I., Didier R., Meinzel R.;  
RT "Characterization of cattle SCO-spondin.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ133488; CAB46239.1;  
DR HSSP; P56682; 1CCV.  
DR InterPro; IPR000359; Cys\_knot.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001545; Glyco\_hormone\_beta.  
DR InterPro; IPR002919; TIL.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR001007; VWFC.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF01826; TIL; 5.  
DR Pfam; PF00090; tsp\_1; 1.  
DR SMART; SM00041; CT; 1.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00068; GHB; 1.  
DR SMART; SM00209; TSPI; 1.  
DR SMART; SM00214; VWC; 2.  
DR PROSITE; PS01225; CTCK\_2; 1.  
DR PROSITE; PS00261; GLYCO\_HORMONE\_BETA\_1; UNKNOWN\_2.  
DR PROSITE; PS50092; TSPI; 1.  
DR PROSITE; PS01208; VWFC; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 1637 AA; 173347 MW; 4C5BAB1DD346C925 CRC64;

Query Match 5.4%; Score 135; DB 6; Length 1637;  
Best Local Similarity 23.4%; Pred. No. 0.0012;  
Matches 59; Conservative 17; Mismatches 70; Indels 106; Gaps 12;

QY 150 SWSVSPDWRAMQWSLSLARANSQDQDYXYSDDSNFLNPPRGWDHTAPGHRFTETK 209  
DB 644 AWS-----RWSAWSPCSRSCGPA--GQSRFRSSTSG-----SW---APECREQSQ 685  
QY 210 DQP-----EYDSTDGEG-----DMSLW 226  
DB 686 SQPCPSQPLCLQGTFRPRSLGDSWLDQCCQCCCTPEGLICEDAEACAGLAWTPSPW 745  
QY 227 SVCSVTGCGNGNQKTRSC-----CYACTA--TESRTCDRPNCPGIEDT----- 267

Db 746 SDPCVSCGGNQVTRVCVASAPRGGSCPLGDPDVQSQRGLWPCPALPDTCSWGPWC 805  
Qy 268 -----FRTAATEVSLAGSEEFNATKLFVDTTSC-----ERWMSCK 304  
Db 806 SRSCGPGGLASASPCCLLAAEPACNSTSP-RLDTQACYAGPCLEFCVWSSWSSWTRCS 864  
Qy 305 SEEL--KKYMHK 314  
Db 865 CEVLVQQRH 876  
RESULT 7  
ID Q95822 PRELIMINARY; PRT; 461 AA.  
AC Q95822, 2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HL01082P.  
GN FAT-SPONDIN.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY060990; AAL28538.1; -  
SQ SEQUENCE 461 AA; 51517 MW; B77D07F41D681B1 CRC64;

Query Match 5.3%; Score 133.5; DB 5; Length 461;  
Best Local Similarity 20.4%; Pred. No. 0.00028;  
Matches 92; Conservative 50; Mismatches 151; Indels 159; Gaps 22;  
Qy 80 PPRPRFRQETGHPSL-----QRDFRSFLDLPNFPDLSKADINGQXPNQVIEV 131  
Db 72 PYNPKSREMTPLAKLYLRREKIVSRNCDDEFLQAL-----QLEV 111  
Qy 132 VGPDPSEADKQHPENKPSWSPDRAWQORSILARA-NSGDODYXYDSTDDSNFL 190  
Db 112 SD--DAE-EQDTRACRVG-----DYSAWSFSCSVSCGKGTMRSRQYLYPAAADQNKCA 162  
Qy 191 NP-----PRGWDHTAPGHRFTFTKQPEYD-----STDGEG----- 221  
Db 163 RLQVAKEMCAIPEACDGA-----QSKDRDDDEGENANSQSLSVSGEGAGLCKTS 216  
Qy 222 DWSLMSVSVTCGNGNQKRTS-----CGYACTATESRTCDRPN-----PG 263  
Db 217 PWSVNSECSACSGIGITRTFVNLGRKRCPH-ITIVKMKMRPCTVEQVELPDQ 275  
Qy 264 IEDTFTATEVSLAGSEEFNATKLFV-----DPTDCERWMSCKSEFLKMYMHKVMNDL 319  
Db 276 CPTOSWSDSPSCSCTCGRVVTRTLLENGPDRESCTQME-----LHQQKECV 326  
Qy 320 PSCPCSYPTVEAYSTADIFDRKDRFKWDAAGPKLEIYKPTARYCIIRSMLSLETT 379  
Db 327 NPIDCHINAEQA-----KD-----ICVQAPDPGPC-RGYMYAYD 361  
Qy 380 LAAQHC-----CYGDNMQLTTRCKGAGTGNLSTERSAELHYKVDVLPWICKGDS- 431  
Db 362 PQNHCHYSFTYGCGRNRNFTDCLNTCNVLRSPYSS-----RVD-QPRACVLSWDSV 416  
Qy 432 -----RYNEARPPNNGKCTE 447

Db 417 WSPCSVSCGVGVSESRRYVVTPEFQNGQPCSK 448  
RESULT 8  
Q9TTS5  
ID Q9TTS5 PRELIMINARY; PRT; 685 AA.  
AC Q9TTS5, 2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SCO-SPONDIN (FRAGMENT).  
GN SCO-SPONDIN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SUBCOMMISSURAL ORGAN;  
RA Gobron S., Creveaux I., Monnerie H., Elbitar F.;  
RT "Characterization of cattle SCO-spondin";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ132106; CAB53759.1; -  
DR InterPro; IPR002919; TIL.  
DR InterPro; IPR000884; TSPL.  
DR Pfam; PF01826; TIL; 2.  
DR Pfam; PF00090; tsp\_1; 6.  
DR SMART; SM00209; TSP1; 6.  
DR PROSITE; PS50092; TSP1; 4.  
DR NON\_TER 1  
FT NON\_TER 685  
SQ SEQUENCE 685 AA; 71389 MW; A003F880E746D5D3 CRC64;  
Query Match 5.3%; Score 133.5; DB 6; Length 685;  
Best Local Similarity 22.6%; Pred. No. 0.00049;  
Matches 78; Conservative 31; Mismatches 137; Indels 99; Gaps 18;  
Qy 146 ENKPSVSPDRAWQORSILARANSQDQXYDSTDDSNFLNPPRGWDHTAGHRT 205  
Db 264 EGQPTWS---PWTWSECSASCPARRKHKRFCTRPGGAPSSMAPPLLSSVPLCPG 319  
Qy 206 FETKQD---PEYDSTDEGDSLMSVSVTCGNGNQKRTSC----- 244  
Db 320 PEAECEPLLPECDRAGGPGWPGFWSCSRSRSGGLRSRACDQPPQGLDYCEGPRA 379  
Qy 245 -GYACTATESRTCDRPNCPGIEDTFTATAEVSLAGSEEFNATKLFVDTDCSERWMSCK 303  
Db 380 QGRACQAL---PCPVTNCTAIE-----GAE-----YSACGPPCPR--SC 413  
Qy 304 KSEFLKMYMHKVMNDLPSPCSYPTVEAYSTADIFDRKDRFKWD-ASGPKEK--LEI 360  
Db 414 DD-----LVHCVHWCQPCYCP-PGQVL--SADGTVHVQPGHCSCDLLTGERHPPGAQL 465  
Qy 361 YKPT-ARYCIIRSMLSLESTL-----AAQHCYGDNMQLITRGKAGTGN 404  
Db 466 AKPDGCMYCTCSEGQLTCTDLPCVPGWGPCWSEWNTACSPQCQ---OTRTRACSCP- 521  
Qy 405 LISTEFAELHYKVDVLPWICKGDSRYNEARPPNNGKCTESP 449  
Db 522 -----APQHGAP-----CPGE---AGEAGAQHRETCASTP 550  
RESULT 9  
ID Q9XZD0 PRELIMINARY; PRT; 763 AA.  
AC Q9XZD0;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE FAT-SPONDIN PROTEIN.  
GN FAT-SPONDIN OR CG6953.

Query Match 5.3%; Score 133.5; DB 5; Length 763;  
Best Local Similarity 20.4%; Pred. No. 0.00057;  
Matches 92; Conservative 50; Mismatches 151; Indels 159; Gaps 22;

QY 80 PPRPRFRQETGHPSL-----QRDFPSFLDLDPNFDLSKADINGXPNQTIVTEV 131  
|| | : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
DB 374 PFYNPKSRMTPLAKLYLRREKIVSRNCDDFEAL-----OLEV 413  
|| | : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
QY 132 VDGPDSEADKDQHPENKPNSVPSPDWRANWORSLSLARA- NSGDQDYXYDSTSDSNNFL 190  
|| | : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
DB 414 SD--DAE-EQDTAEARGV-----DYSANSPCSVSCKGIRMSROYLYPAAADQNKA 464  
|| | : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
QY 191 NP-----PRGMHTAPGHRTFETKDKPEYD-----STDGEG----- 221  
|| | : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
DB 465 RQLVAKEMCVAAIPECADGPA-----QSKDRDDDEGENLANSSOLVGSNGEGAGLCKTS 518  
|| | : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
QY 222 DWSLWSVCSTCGNGNQKRTRS-----CGYACTATESRCDRNC-----PG 263  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
DB 519 FWSWSECSASCCIGITMTRTTFVNHLGRKCRPH-ITIVEKNKMREDPCITYEQVELPPQ 577  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
QY 264 IEDTFTATAEVSLLAGSEFNATKLFV----DTDSCERWMSCKSEFLKKYMHKVMDL 319  
: : : : : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
DB 578 CPTSQSWDSWSPCSTCGRGVTIRULLLENGDPDKESCOTRME-----LHQKECV 628  
: : : : : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
QY 320 PSCPSPYTEVAYSTADIFDIRIKRKOPRWKADSGPKKELEYPTARYCYRSMLSLESTT 379  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
DB 629 NPTDCHINAEQA-----KD-----ICVQAPDPGPC-RGTVMRYAYD 663  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
QY 380 LAAQHC-----CYGDNMOLLTRKGAGTPNLISTEFSAEELHYKVDVLPIWICKGDWS- 431  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
DB 664 PQNHOCYSFYGGCGRRNNFLFENDCLNTCNVLRSPYS-----RVD-QPRACVLSDWSV 718  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
QY 432 -----RYNEARPPNNGKCTE 447  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
DB 719 WSPCSVSGVGVSERSYVTEPQNGQPCK 750

RESULT 10 function  
Q23832 PRELIMINARY; PRG; 660 AA.

ID Q23832 AC Q23832; 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE MICRONEMAL TRAP-C1 PROTEIN HOMOLOG (FRAGMENT).  
TRAP-C1.  
OS Cryptosporidium wrairi.  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;  
OC Cryptosporidiidae; Cryptosporidium.  
OX NCBI\_Taxid=35083;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=GUINEA PIG ISOLATE;  
RX MEDLINE=98234217; PubMed=9574918;  
RA Spano F., Putignani L., Naiza S., Puri C., Wright S., Crisanti A.;  
RT "Molecular cloning and expression analysis of a Cryptosporidium parvum  
RL Mol. Biochem. Parasitol. 92:147-162(1998).  
RL EMBL; U42213; AAC48313.1;  
DR InterPro: IPR000884; TSPl.  
DR Pfam: PF00090; tsp.1; 5.  
DR SMART: SM00209; TSP1; 6.  
DR PROSITE; PS50092; TSPl; 5.  
DR NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 660 AA; 72812 MW; FFF45490B8812091 CRC64;

Query Match 5.2%; Score 131.5; DB 5; Length 660;  
Best Local Similarity 22.0%; Pred. No. 0.00073;  
Matches 69; Conservative 33; Mismatches 116; Indels 95; Gaps 15;

```
QY 158 WRAMORSLSLARANGSDQDYXYDSTSDSNFLNPPRGWDHTAPGHRTEET--KQOPEYD 215
Db 352 WSEWSDCSTG---CGEGNRTRTETSKP-----PLNGDESTCPELIAKESCNK-VECP 401
QY 216 STDGE-GDWSLWSVCSVTCGNGNQKTRSC-GYACT--ATESRTCDRPNCPGIEDTFRTA 271
Db 402 NIOCELGEWSSWSPCVTCGSGTTSRNRVKGNCTELPESKCKLANC-----451
QY 272 ATEVSLLAGSEENATKLFEDVTDSCERWMSKSE----FLAKYMHKVMNDLPSCPSYP 327
Db 452 -----GDNSASCTAVMSVSE-----WSACEKCDQGLVRRY-----483
QY 328 TEVAYSPADIFDRIRKRDKFRKWDASGPKELEYKPTARYICRSMLSLESTTLAAQHOCY 387
Db 484 RDFDFSKIGVGVV-----PPGKSEBQNKVREIC-----KDTPTLEEBPCTS 525
QY 388 GDNQMLITRCKGAGTPNLISTERSA-----ELHYKVDVLPWLICKGDWSRYNEARP 438
Db 526 GVTC-----TPGCKYTEWSAWSSDCSGSOTRDRVVTPEPGIIDAICOSSKDTRS 575
QY 439 PNNGOKCTESPSD 451
Db 576 CSRPEGCTETTPD 588

RESULT 11
Q9U8J9 PRELIMINARY; PRT; 765 AA.
AC Q9U8J9
AC Q9U8J9
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE THROMBOSPONDIN-RELATED ADHESIVE PROTEIN HOMOLOG.
GN MTC2.
OS Neospora caninum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
OC Neospora.
OX NCBI_TaxID=29176;
RN [1]
RX MEDLINE=20183852; PubMed=10717300;
RA Lovett J.I., Howe D.K., Sibley L.D.;
RT "Molecular characterization of a thrombospondin-related anonymous
RT protein homologue in Neospora caninum.";
RL Mol. Biochem. Parasitol. 107:33-43(2000).
DR EMBL; AF061273; AAF01565.1; -
DR InterPro; IPR001969; Asp.protease.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00209; TSPI; 6.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS00092; TSPI; 6.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 765 AA; 82880 MW; 9727838CC1FE8CD4 CRC64;

Query Match 5.08; Score 127; DB 5; Length 765;
Best Local Similarity 21.88; Pred. No. 0.0025;
Matches 60; Conservative 34; Mismatches 81; Indels 100; Gaps 14;

QY 135 PDSEADKQHPENKPSVSP--SPDWRAMWORSLSLARANGSDQDYXYDSTSDSNFLNP 192
Db 382 PPVEKSQTEQAQFAPCVVIPPTEPSAW---SECTATCGGGTRH-----423
QY 193 PRGWDHTAPGHRTE--TKDOPE-----YDSTGE-----GDWSLWSVCSV 231
Db 424 -RSRNLPGPGRSADQNTQEPKPSNPWPFGDLHEQSCNNSPCPINATCGDFGEWSECSV 482
QY 232 TCGNGNQKTR-----SC-GYACTATESRTCDRPNCPGIED-----266
```

```
Db 483 SCGGLSORSRDPWNNDQHGKSCMQQYPNGHTEKRSNAQPCPVDPEPGDWEWBCN 542
QY 267 -----TFTTAATEVSVLLAGSE---EFNAT-----KLFEVDTDSC-----297
Db 543 VTCGGGTRRRGRSVILPQYGRSIVEBQNSKLPENKILLIVETETCSLPCDASCTFPW 602
QY 298 ERWMS---CKSEFLKYYMHKVMN-DLPSCPSYPT 328
Db 603 SDWSSCEQCEAGTGTQYRNSAVKFDYRNKPCDFT 637

RESULT 12
Q9V692 PRELIMINARY; PRT; 873 AA.
AC Q9V692
AC Q9V692
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CG1739 PROTEIN (GH02025P).
GN CG17739.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfaffler B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RX SEQUENCE FROM N.A.
RC STRAIN=X, CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J., R.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
```

RA	Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA	Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA	Yu C., Lewis S.E., Rubin G.M., Ceiniker S.
RL	Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AE003822; AAF58536.2; -
DR	EMBL; AY058267; AAL13496.1; -
DR	HSPB; P00980; IDTX.
DR	FlyBase; FBgn0033710; CG17739.
DR	InterPro; IPR002223; Kunitz_BPTI.
DR	InterPro; IPR002861; Reeler.
DR	InterPro; IPR000884; TSPI.
DR	Pfam; PF00014; Kunitz_BPTI; 2.
DR	Pfam; PF02014; Reeler; 1.
DR	Pfam; PF00090; tsp_1; 7.
DR	SMART; SM00131; KU; 1.
DR	SMART; SM00209; TSP1; 5.
DR	PROSITE; PS02279; BPTI_KUNITZ_2; 1.
DR	PROSITE; PS50092; TSPI; 5.
DR	Serine protease inhibitor.
KW	SEQUENCE 873 AA; 98281 MW; 62BA97AAFBA42CD1 CRC64;
SQ	

  

Query Match	5.0%; Score 125.5; DB 5; Length 873;
Best Local Similarity	19.4%; Pred. No. 0.0041;
Matches	75; Conservative 31; Mismatches 132; Indels 149; Gaps

  

Qy	140	DKDQHNPKNPSVPSDW-----RAMQRSLSLARNSGDQDYXDSTSD	186
	:	: : : : :	:
Dd	287	DSNNHLLSLYSWDPSPDIWGVSGLELCPLNCWSWE-----NKVHNLPWDAGTDS	338
	:	: : : : :	:
Qy	187	-----SNLNPPRG--WDHTAPGHRFTETKD-----QPEYDS	216
	:	: : : : :	:
Dd	339	GPSYMSADQPVPDDVVRRYIKSNFPNDPRSFPFDPTGAQMKLATLIHNRRLRYEKNCS	398
	:	: : : : :	:
Qy	217	TDGE-----GDWSLWSCSVTCGNQKRTS-----CGYACTATESPTCD	257
	:	: : : : :	:
Dd	399	SDSEQVPEGCATNWSRWDECTTKCGPKQYRIREFKPNPALASRHRCNNA--LRSEKNCV	456
	:	: : : : :	:
Qy	258	RPNCPGI-EDTFRTAATFEVLLAGSEFNATKLFVDVTOSC-----ERWMSCKSEFLKKY	311
	:	: : : : :	:
Dd	457	GHKCAGNEETAEGGEVEVAPSGSS-----DDQCGLSDHWSEWSSC-----	498
	:	: : : : :	:
Qy	312	MKVWNLDLPSCPSYPEVAYSTADIFDIRIKRDFRKWDASGPKELEYKPTARYCIIR-	370
	:	: : : : :	:
Dd	499	-----TVTCGTGEM-----TRSRHLNKKAKKKCKQA	525
	:	: : : : :	:
Qy	371	SMLSLETTTLAAQHCYCDNWQLITRCKGAGTP-----NLISTEPSAEHL	415
	:	: : : : :	:
Dd	526	SRARLHETKICEAMECCGD----IENEGGAGEPEEQAGDGGSAAEKRSIFRNFSYSQ	581
	:	: : : : :	:
Qy	416	YKVDVLPIWICKGDWSRYNEARPNNNG	442
	:	: : : : :	:
Dd	582	HREDYIPVCGVTPWSDFPSCMGPCGG	608
	:	: : : : :	:

  

RESULT	13
ID	Q19791 PRELIMINARY; PRT; 2165 AA.
AC	Q19791; Q27524;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	F25H8.3 PROTEIN.
GN	F25H8.3.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_TaxID=6239;
RN	[1]
RN	SEQUENCE FROM N.A.
RP	Gajadasy S.;
RL	Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN	[2]





Qy	361	- - - - - YKTFARY -- C I R S M L S E S T T L A A Q H C Y G D N M O L - - - - - I T R G K G A C T P N	404
Db	3535	L D D C F E A D G G F G P W S P W G P C S R S C G G L G T R T R S R Q - - C V L T M P T I S E L P V C P G P G C G A G N	3592
Qy	405	L I S T F S A E L U H Y K V D L P W I N I C K - - - - - G D S R Y N E A R P P N - N G Q K T E S P S D E D Y I K Q F	458
Db	3593	C S W T S N A - - - - - P W E P C S C G V Q Q R R L A Y R P P G P G G H W C - - - - - P N I L T A Y	3636
Qy	459	Q E A R	462
Db	3637	Q E R R	3640

Search completed: April 25, 2002, 17:26:33  
Job time: 9600 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: April 25, 2002, 13:59:38 ; Search time 44.51 Seconds  
(without alignments)  
1001.695 Million cell updates/sec

Title: US-09-863-824-2  
Perfect score: 2517  
Sequence: 1 MVRLAELLGLLLTLH.....CTSPSDEYIKOFQAREY 464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143	5.7	1572	2 T00027	brain-specific ang
2	137.5	5.5	1584	2 T00026	brain-specific ang
3	134.5	5.3	984	2 T00326	hypothetical prote
4	134.5	5.3	1522	2 T00028	brain-specific ang
5	131	5.2	437	2 S05478	properdin - mouse
6	127.5	5.1	1170	2 A40558	thrombospondin 1 p
7	127	5.0	1170	1 TSHUP1	thrombospondin 1 p
8	124	4.9	1205	2 T18517	procollagen N-endo
9	124	4.9	2165	2 T21371	hypothetical prote
10	122	4.8	837	2 T00355	hypothetical prote
11	120.5	4.8	1172	2 A42587	thrombospondin 2 p
12	119	4.7	1178	1 A39804	thrombospondin pre
13	119	4.7	1444	2 T18856	angiogenesis inhib
14	118	4.7	957	2 T15976	hypothetical prote
15	117.5	4.7	807	2 A38152	F-spondin - rat
16	115	4.6	788	2 T25061	hypothetical prote
17	115	4.6	1172	1 TSHUP2	thrombospondin 2 p
18	114	4.5	3562	2 A47171	chondroitin sulfat
19	111.5	4.4	1059	2 T22545	hypothetical prote
20	111	4.4	469	1 S29126	properdin precurs
21	109.5	4.4	2403	2 A59386	sanko - human
22	106.5	4.2	803	2 A47723	F-spondin precurs
23	104.5	4.2	559	2 S04531	thrombospondin-rel
24	104	4.1	513	1 I58311	HMG-box containing
25	104	4.1	610	2 T16761	hypothetical prote
26	104	4.1	660	2 T39964	hypothetical colle
27	104	4.1	724	2 A48569	antigen Em100 - Ei
28	104	4.1	1074	2 JC5928	semaphorin F precu
29	104	4.1	1360	2 T33922	hypothetical prote

30 104 4.1 1558 2 C89114 protein C37c3.6a l  
31 104 4.1 2167 2 T34395 hypothetical prote  
32 103 4.1 2098 2 T18397 protein Ctrp - mal  
33 102 4.1 2026 1 OVBY adenylate cyclase  
34 101 4.0 1340 2 A39808 proteoglycan core  
35 101 4.0 2327 2 T24630 aggrecan - bovine  
36 100.5 4.0 654 2 T29247 hypothetical prote  
37 100.5 4.0 1643 2 T14274 versican precursor  
38 100 4.0 805 2 T34212 hypothetical prote  
39 99.5 4.0 440 2 T24232 hypothetical prote  
40 99.5 4.0 934 1 A34372 complement C6 prec  
41 99 3.9 2058 2 A59267 myosin x - human  
42 98.5 3.9 977 2 I52657 seizure-related pr  
43 97.5 3.9 550 2 T47158 hypothetical prote  
44 97.5 3.9 574 2 A46283 sporozoite surface  
45 97.5 3.9 753 2 S62411 probable guanine n

## ALIGNMENTS

RESULT 1  
T00027  
brain-specific angiogenesis inhibitor 2 - human  
N:Alternate names: BAI2 protein  
C:Species: Homo sapiens (man)  
C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
C:Accession: T00027  
R:Shiratsuchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.  
Cytogetet. Cell Genet. 79, 103-108, 1997  
A:Title: Cloning and characterization of BAI2 and BAI3, novel genes homologous to bra  
A:Reference number: Z14066; MUID:98194217  
A:Accession: T00027  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-1572 <SHI>  
A:Cross-references: EMBL:AB005298; NID:g3021698; PIDN:BAA25362.1; PID:g3021699  
A:Experimental source: brain  
C:Genetics:  
A:Gene: GDB:BAI2  
A:Cross-references: GDB:9838089; OMIM:602683  
A:Map position: lp35-1p35

Query Match 5.7%; Score 143; DB 2: Length 1572;  
Best Local Similarity 19.5%; Pred No. 0.0068;  
Matches 89; Conservative 47; Mismatches 161; Indels 160; Gaps 19;  
QY 114 SKADINGQXPNIQVTVIEVVDGPDSEADKQHPENKPSW--SVSPSP-----DW 158  
Db 248 AEADLHSGSSNDLFTTEMRYGEEPE---EPPKVKTOWPRSADEPGLYMAQTGDPAAEW 303  
QY 159 RAWQBSLSLARNSGDQDYXYDSTSDSNFL-----NPPRGWDHTAGHRTFE---- 207  
Db 304 SPMSVCSLT---CGQGLQVTRSCVSPSYGTLCSGRLRTRPCNNSATCPVHGWEWGS 360  
QY 208 -----TKDQPEYDSTDGEG-----DMSLSVCS 230  
Db 361 WSLCSRSCGRSRRMRTCVPPQHGKACEGPELQTKLSMAACPVEGGWLEHGWGPGPCS 420  
QY 231 VTGNGNQKTRSCGYA-----CTA--TESRTCDRNPCEGIEDTFRTAATEVSLLAGSE 282  
Db 421 TSCANGTQQRSKCSVAGPAWATCTGALTDTRECSNLECCPATDSKW-----G 467  
QY 283 ENATYKLF--VDTDCERHMSCKSEFLKMYHKVNDLPCSPCSYPTVEAYSTAD---- 336  
Db 468 PMAWSLCSKTCDTGQWRFRMCQATGTQY-----PCEGTGEVKEPCSEKRCP 516  
QY 337 IEDRIKKDF---RWKDSGPKLEIKYKPTA---RYCIRS----- 371  
Db 517 AFHEMCRDEYVLMTWKAAAGETIYNKPPNASSGSASRRCLLSAOGVAVWGLPSFARCI 576  
QY 372 -----MLSLESTTLAAQHCCYGDNNQLITRKGAGTNPILIS--TEFSAELHYKVDVLP 422

Db 577 SHEYRYLSLREHLAKGORMLAGESOVVR-----SLQELLARRTYSGDLLESVDIL- 631  
 QY 423 WLICKGWSRYNEARPPNNGOKCTESDDEYIKQFQ 459  
 Db 632 -----RNVDTFRATYVPSADVDQRFQ 655

RESULT 2  
 T00026  
 brain-specific angiogenesis inhibitor 1 - human  
 N:Alternate names: BAI1 protein  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 12-Feb-1999  
 C:Accession: T00026  
 R:Nishimori, H.; Shiratsuchi, T.; Urano, T.; Kimura, Y.; Kiyono, K.; Tatsumi, K.; Yoshida  
 submitted to the EMBL Data Library, June 1997  
 A:Reference number: Z14064  
 A:Accession: T00026  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1584 <NIS>  
 A:Cross-references: EMBL:AB005297; NID:d1175078; PID:d1024528  
 A:Experimental source: brain  
 C:Genetics:  
 A:Gene: GDB:BAI1  
 A:Cross-references: GDB:9838088; OMIM:602682  
 A:Map position: 8q24-8q24  
 C:Superfamily: thrombospondin type 1 repeat homology  
 F:408-462/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 5.5%; Score 137.5; DB 2; Length 1584;  
 Best Local Similarity 21.8%; Pred. No. 0.019;  
 Matches 119; Conservative 50; Mismatches 177; Indels 201; Gaps 30;  
 QY 2 VRLAELLLLLGLLLTLHTVLRSGAAGDPA--GNAQAQLNNLNVS 53  
 Db 11 WVLAPLILL--LLLGRRAAAGADAGPPECATLVQGFYFAA--VFP 62  
 QY 54 DTTSETSFSLKEAPREH-LDHOAAHQPPRFRQETGHPSLQDFP-RSFL-----L 105  
 Db 63 ANASRCWTLRPPDPRRTYLYMKVAKAPV-----CSGGRVRYQDFSLFLESTRYL 115  
 QY 106 DLPNFPD-LSKADINGXPNIOVTTEVDGPDSEADKQHPEN---KPSWSVSP--DW- 158  
 Db 116 GVSEFDEVLRCDPSAPLAFQASKQFL-----QMRQPPQHDGLRPAGPPGPTDDES 170  
 QY 159 -----RA-----WQRLSLARAN-----SGDDYXYDSTDD 186  
 Db 171 VEYLVVGNRNPRAAQLCRLWDLACLAGRSHPGIMQTPCACLGGAG----- 221  
 QY 187 SNFLNP-----PRG-----WDHTAQRH---TFETKQPEYDSTDEGDSLWSVCSVTC 233  
 Db 222 ---GPAAGPLAPRGDVCLURDAVAGPENCLSLTQDRGHGATGKWLWSLWGECTDRC 277  
 QY 234 GNGNQRTRSC-----GYACTAT--ESRTCDRPNCPGIEDTFRATAAEVSLA----- 279  
 Db 278 GGLQTRTRTCLPAPGVEGGCGEVLLEGRCQNRAC---GPAGRTSSRSQSLRSTDARR 334  
 QY 280 ----GSE----EFNATKLFVETDSCERWMSCKSEFLKYMVKYVMDLPSCPSPTEVA 331  
 Db 335 REELGDELQFGFPAPQTDGPAAEWSPWSVCS-----TCG----- 371  
 QY 332 YSTADIFDRKDRFKWDASGPKEKLEIYKPTARYCIRMSLSLE-STTLAAQHCCYGN 390  
 Db 372 -----EGWQTRFCVSSVSTQCSGPIRQRLC---N 401  
 QY 391 MQLITRGKAGTPNLISTEFSAEHLKYVDVLPWIIIC-----RGDWSRYNEARPPNNGQRC 445  
 Db 402 NSACVPHGAW-----DEWS-----PWSLCSSTCGRFRDTRCTRPQPGGPNP 445  
 QY 446 TESPSE 452

Db 446 CEGPEKQ 452  
 RESULT 3  
 T00326  
 hypothetical protein KIAA0550 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00326  
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,  
 DNA Res. 5, 31-39, 1998  
 A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp  
 A:Reference number: Z14086; MUID:98290545  
 A:Accession: T00326  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-984 <NAG>  
 A:Cross-references: EMBL:AB011122; NID:g3043623; PIDN:BAA25476.1; PID:g3043624  
 A:Experimental source: brain  
 C:Genetics:  
 A:Note: KIAA0550  
 C:Superfamily: thrombospondin type 1 repeat homology  
 F:344-398/Domain: thrombospondin type 1 repeat homology <THR3>  
 QY 62 SLSKEA---PREHL-----DHQAAHQPPRFRQETGHPSLQDFPFRSFLD-----LPN 109  
 Db 241 NLTREAKRPKEEFMGMDHTIKSQ---RPRSVHEKRVPPQEQADA--FMAQTGESGVEE 296  
 QY 110 FPDLSKADIN--GOXPNIQVTEV-----VDGPDSEADKQHPENKPSWSVSPDWRRAW 162  
 Db 297 WSOVSTCVTCGQSQVTRTCVSPYTHCSGRLRESRVNCNTALCPVHG--WEWS 352  
 QY 163 QRLSLARANSGDDYXYDSTSDSNFLNPPRGWDHTAPGHRFTETKQDEYDSTDGE-G 221  
 Db 353 PWSLCSFTCGRQTRTRSC-----PPQYGRGPCGEPETHHKPCNIALCPVDGQWQ 404  
 QY 222 DWSLWSVSVTCGNGNQKTRSC-----GYACTA--TESRTCDRPNCPGIEDTFRATA 273  
 Db 405 EMSWSQSVTCNSGTOQRSQCTAAAHGSGECRGPWAESRECYNPEC----- 452  
 QY 274 EVSLLAGSEEFNATKLFVETDSDCE-----RWMSCKSEFLK--KYMVKYVMDLPSCP---C 324  
 Db 453 -----TANGQWQWGHWSGCKSCDGGWERRIRTCQGAIVTCQCEGTGEVRRCSQRC 507  
 QY 325 SYPTVEA---YSTADIFDRIRKKDFRWK-----DASGPKEK---LEIY-----KPTARYC 368  
 Db 508 PAPYEICPEDYLMVMVKRTPDAGDLAFNQCLNATGTTSSRCSLSLHGVAFWEQPSFARC 567  
 QY 369 IRS-----MLESSTTLAAQHCCYGNMQLITR 396  
 Db 568 TSNEYRHQHSIKEHLAKQRLAGDGMQSVTK 600  
 RESULT 4  
 T00028  
 brain-specific angiogenesis inhibitor 3 - human  
 N:Alternate names: BAI3 protein  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
 C:Accession: T00028  
 R:Shiratsuchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.  
 Cyogenet. Cell Genet. 79, 103-108, 1997  
 A:Title: Cloning and characterization of BAI2 and BAI3, novel genes homologous to bra  
 A:Reference number: Z14086; MUID:98194217  
 A:Accession: T00028  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A:Residues: 1-1522 <SHI>  
A:Cross-references: EMBL:AB005299; NID:g3021700; PIDN:BAA25363.1; PID:g3021701  
A:Experimental source: brain  
C:Genetics:  
A:Gene: GDB:BAT3  
A:Cross-references: GDB:9838090; OMIM:602684  
A:Map position: 6q12-6q12  
C:Superfamily: thrombospondin type 1 repeat homology  
F:344-398/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 5.3%; Score 134.5; DB 2; Length 1522;  
Best Local Similarity 21.4%; Pred. No. 0.031;  
Matches 84; Conservative 48; Mismatches 170; Indels 91; Gaps 21;  
  
Qy 62 SLSKEA---PREHL-----DHQAAHPPFRFRQETGHPSLQRDPFRSFLLD-----LPN 109  
Db 241 NLTREAKRPPEFGMGMDHIKSO---RPRSVHEKRVPEQADAQK-FMAQTGSGVEE 296  
  
Qy 110 FPDLSKADIN-GQXPNIQVTIEV-----VDGPDSEADKQHPENKPSWSVSPDWRAMW 162  
Db 297 WSNWSTCVTCQGSQVTRTCVSPYTHCSGPLRESRVNNTALCPVHGV---WEWS 352  
  
Qy 163 QRSLSLARANGSDQDYXYDSTSDSNFLNPPRGWDHTAPGHTFTETKDOPEYDSTDGE-G 221  
Db 353 PWSLCSFTCGRQRTTRRSCT-----PPQYGRPCGEPETHHKPCNIALCPVDGQWQ 404  
  
Qy 222 DWSLWSVSVTCGNGNOKRTSC-----GYACTA--TESRTCDRPNCPGIEDTFTTAAT 273  
Db 405 EWSWSQCSVTCNSGTQQRQROCTAAAHGGSECRPWAESREYNPEC----- 452  
  
Qy 274 EVSLLAGSEFNATKLFVDVTDSC-----RWMSCSKSEFLK-KYMHKVMNDLPSCP---C 324  
Db 453 -----TANGQWNQWCHWSGCSKSGDGGWERRIRTCQGAIVTQCGQCEGVEVRRCSEQR 507  
  
Qy 325 SYPTVA---YSTADIFRIKKDRWK-----DASGPKEK---LEIY-----KPTARYC 368  
Db 508 PAPYEICPEDYLMVMVKRTTPAGDLAFNOCPLNATGTTSRCSLSLHGVAFWEQPSFARC 567  
  
Qy 369 IRS-----MLSLESTTLAAHCYCGDNMOLITR 396  
Db 568 ISNEVRHLQHSIKEHLAKQRMGLAGDGMGSQVTK 600

RESULT 5  
S05478  
properdin - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 17-Nov-2000  
C:Accession: S05478  
R:Goundis, D.; Reid, K.B.M.  
Nature 335, 82-85, 1988  
A:Title: Properdin, the terminal complement components, thrombospondin and the circumspon  
A:Reference number: S05478; MUID:88318954  
A:Accession: S05478  
A:Molecule type: mRNA  
A:Residues: 1-437 <GOU>  
A:Cross-references: EMBL:X12905; NID:g53786; PIDN:CAA31389.1; PID:g53787  
C:Complex: a mixture of homodimers, homotrimers and homotetramers  
C:Function:  
A:Description: protects C3 convertase (C3bb) from rapid inactivation  
A:Pathway: complement alternate pathway  
C:Superfamily: human properdin precursor; thrombospondin type 1 repeat homology  
F:45-97/Domain: thrombospondin type 1 repeat homology <THR1>  
F:104-160/Domain: thrombospondin type 1 repeat homology <THR2>  
F:161-224/Domain: thrombospondin type 1 repeat homology <THR3>  
F:225-282/Domain: thrombospondin type 1 repeat homology <THR4>  
F:283-345/Domain: thrombospondin type 1 repeat homology <THR5>  
F:348-408/Domain: thrombospondin type 1 repeat homology <THR6>  
F:525-55108,111,114,165,168,229,232,290,293,350,353,356/Modified site: 2'-mannosyl-trypt  
F:366,396/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.2%; Score 131; DB 2; Length 437;  
Best Local Similarity 26.5%; Pred. No. 0.012;  
Matches 53; Conservative 16; Mismatches 79; Indels 52; Gaps 11;  
  
Qy 155 SPDRAWNQRSLSLARANGSDQDYXYDSTSDSNFLNPPRGW---DHTAPGHTTFB---T 208  
Db 46 SPQSAWSLWGPCSVTCSEGSQLRHRRVCG-----RGQCSENVAPGTLLEWQAC 96  
  
Qy 209 KDQPEYDSTDGEGDWSLWSVTCGNGNQKRTR-----SCGYAC--TATESRTCD-R 258  
Db 97 EDQPCPEMGWSEWGPWGPCSVTCSKGQIRQVCDNPAPKCGGHCPCGAQSQACDTQ 156  
  
Qy 259 PNCPGIEDTFTTAAT-----EVSLLAGSEFNATKLFVDVTDSCER-----WMSC 303  
Db 157 KTCP-----THGAWASWGPSPRSGSLCGAQEPKETR-----SRSCSAPASHQPPGKPC 207  
  
Qy 304 KSEFLKKYMHKVMNDLPSCP 323  
Db 208 SG---PAYEHKACSGLPSCP 224

RESULT 6

A40558  
thrombospondin 1 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 20-Aug-1999  
C:Accession: A40558; A37905; B42587; S68787  
R:Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.  
Genomics 11, 587-600, 1991  
A:Title: Characterization of the murine thrombospondin gene.  
A:Reference number: A40558; MUID:92128941  
A:Accession: A40558  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1170 <LAW>  
A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454;  
GB:M62465; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:g511867; PIDN:AAAS  
R:Bornstein, P.; Alfli, D.; Devarayalu, S.; Framson, P.; Li, P.  
J. Biol. Chem. 265, 16691-16698, 1990  
A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role  
A:Reference number: A37905; MUID:90375546  
A:Accession: A37905  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-490 <BOR>  
A:Cross-references: GB:J05605; GB:J05606; NID:g201991; PIDN:AAA40431.1; PID:g554390  
R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seidin, M.F.; Dixit, V.M.  
J. Biol. Chem. 267, 3274-3281, 1992  
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce  
A:Reference number: A42587; MUID:92147683  
A:Accession: B42587  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1152, 'p', 1154-1170 <LAH>  
A:Cross-references: GB:M87276  
A:Note: sequence extracted from NCBI backbone (NCBIP:81501)  
R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.  
FEBS Lett. 387, 36-41, 1996  
A:Title: Expression and initial characterization of recombinant mouse thrombospondin  
A:Reference number: S68787; MUID:96234006  
A:Accession: S68787  
A:Molecule type: protein  
A:Residues: 19-26, 'x', 28-37 <CHE>  
C:Complex: homotrimer, disulfide linked  
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;  
F:1-18/Domain: calcium binding; glycoprotein; homotrimer  
F:19-1170/Product: thrombospondin 1 #status predicted <WAT>  
F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>  
F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>  
F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>  
F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>

F:551-586/Domain: EGF homology <EGF>  
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.18; Score 127.5; DB 2; Length 1170;  
Best Local Similarity 23.3%; Pred. No. 0.08;  
Matches 41; Conservative 17; Mismatches 53; Indels 65; Gaps 8;

QY 151 WSPSPD--WRAWQRLSLARANGDOD--YYXDSTSD-----DSNF 189

DB 373 WPSDADGWSFWSWSSCSATCGNGIQGRSCDNLNRCEGSSVQTRTCHIQECKRF 432

QY 190 LNPFGWHTAPGHRFTET-----KQPEYDSTDGEGD----- 222

DB 433 -KQDGGWHSWSPSSCSVTCGDGVITRILNCSLSPQMGKPCGEARETRACKKDACPI 491

QY 223 -----WSLWSVCSVTCGNGNQKTRTSC-----GYACT--ATESRTCDRPNCP 262

DB 492 NGGWGWPSPWDICSVTCGGVQRRSLCNPTQFGGKDCVGVDTENQVCKQDCP 547

# RESULT 7

TSHUPI

C:Species: Homo sapiens (man)

C:Date: 23-Aug-1987 #sequence\_revision 03-Aug-1995 #text\_change 17-Nov-2000

C:Accession: A26155; A34274; A30140; A25812; A05172; A42927

R:Lawler, J.; Hynes, R.O.

J. Cell Biol. 103, 1635-1648, 1986

A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple

A:Reference number: A26155; MUID:87057617

A:Accession: A26155

A:Molecule type: mRNA

A:Residues: 1-1170 <LAW>

A:Cross-references: GB:X04665; NID:g37137; PIDN:CAA28370.1; PID:g37138

A:Note: parts of this sequence, including the amino end of the mature protein, were dete

R:Lahterty, C.D.; Gierman, T.M.; Dixit, V.M.

J. Biol. Chem. 264, 11222-11227, 1989

A:Title: Characterization of the promoter region of the human thrombospondin gene. DNA s

A:Reference number: A34274; MUID:89291870

A:Accession: A34274

A:Molecule type: DNA

A:Residues: 1-166 <LAH>

A:Cross-references: GB:J04835

R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwein,

J. Cell Biol. 108, 729-736, 1989

A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in th

A:Reference number: A30140; MUID:89139590

A:Accession: A30140

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>

A:Cross-references: EMBL:X14787; NID:g37464; PIDN:CAA32889.1; PID:g37465

A:Note: parts of this sequence, including the amino end of the mature protein, were dete

R:Kobayashi, S.; Eden-McDutchman, F.; Framson, P.; Bornstein, P.

Biochemistry 25, 8418-8425, 1986

A:Title: Partial amino acid sequence of human thrombospondin as determined by analysis o

A:Reference number: A25812; MUID:87157592

A:Accession: A25812

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-397 <KOB>

A:Cross-references: GB:M25631; NID:g538353; PIDN:AAA36741.1; PID:g538354

R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986

A:Reference number: A05172; MUID:86287276

A:Accession: A05172

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>

A:Cross-references: GB:M14326; NID:g340005; PIDN:AAA61237.1; PID:g553801

A:Note: parts of this sequence, including the amino end of the mature protein, were dete

R:Sun, X.; Skorstengaard, K.; Mosher, D.F.

J. Cell Biol. 118, 693-701, 1992

A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin.

A:Reference number: A42927; MUID:92348511

A:Accession: A42927  
A:Molecule type: protein  
A:Residues: 987-1003 <SUN>  
A:Note: Cys-992 is shown to have a free sulphydryl

C:Genetics:

A:Gene: GDB:THBS1; TSP1; TSP

A:Cross-references: GDB:120438; OMIM:188060

A:Map position: 15q15-15q15

A:Introns: 23/1

A:Note: the list of introns may be incomplete

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1170/Product: thrombospondin 1 #status predicted <MAT>

F:317-375/Domain: von Willebrand factor type C repeat homology <WVC>

F:378-429/Domain: thrombospondin type 1 repeat homology <THRI>

F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>

F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>

F:551-586/Domain: EGF homology <EGF1>

F:650-689/Domain: EGF homology <EGF2>

F:926-928/Region: cell attachment (R-G-D) motif

F:171-232/Disulfide bonds: #status predicted

F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:270,274/Disulfide bonds: interchain #status predicted

F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 5.08; Score 127; DB 1; Length 1170;

Best Local Similarity 44.4%; Pred. No. 0.087;

Matches 24; Conservative 7; Mismatches 13; Indels 10; Gaps 2;

QY 219 GEGDWSLWSVCSVTCGNGNQKTRTSC-----GYACT--ATESRTCDRPNCP 262

DB 494 GGGPWPMDICSVTCGGVQRRSLCNPTQFGGKDCVGVDTENQICNKQDCP 547

# RESULT 8

T18517

procollagen N-endopeptidase (EC 3.4.24.14) I - bovine

A:Alternate names: procollagen N-proteinase

C:Species: Bos primigenius taurus (cattle)

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T18517

R:Collige, A.; Nusgens, B.V.; Lapiere, C.M.

submitted to the EMBL Data Library, February 1996

A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.

A:Reference number: Z18941

A:Accession: T18517

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1205 <COL>

A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1

A:Experimental source: skin

C:Genetics:

A:Gene: PC I-NP

C:Function:

A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior

C:Keywords: hydrolase; metalloproteinase

Query Match 4.9%; Score 124; DB 2; Length 1205;

Best Local Similarity 20.9%; Pred. No. 0.16;

Matches 64; Conservative 26; Mismatches 84; Indels 132; Gaps 15;

QY 192 PPRGWDHTAPGHRTEF-----TKDQPEYDSTDGEGDWSLWS---VCSVTCGNGNQKTRR 242

DB 526 PPLDGTMTCAPGKCHRCGHCINWTPDLTKRD-----GNMGAWSPFGCSSTCTGTVKFRFR 580

QY 243 SC-----GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEFNATKLFVEV 292

Db 581 QCONPHANGRTCSGLAYDFOLCNSQDCPDALADFR----- 617  
Qy 293 DTDSCERWMSKSEFLAKYMHKVMNDLPSCSPYPTVAYSTADIFDIKRKDFRW--- 348  
Db 618 -EEOCRW----- 641  
Qy 349 -KDSAGPKLEIY---KPTARYCIRSMLSLESTLLAAHCYCGDNMQLITRG--KGAGT 402  
Db 642 HRDA---KERCHLYCESKETG-----EVSMSKRMVHDGTRCSYKDAFSLCVRGDCRKVGC 693  
Qy 403 PNLISTESFAB-----LHYKVDVLPWIIKGDMSRYNEARPPNNGOKCTSPDSED 453  
Db 694 DGVIGSSKQEDKCGVCGDGNSHCKV-----VRGTFSR-----SPKKLG 731  
Qy 454 YIKOFQ 459  
Db 732 YIKMFE 737  
RESULT 9  
T21371  
hypothetical protein F25H8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T21371; T24896  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19413  
A:Accession: T21371  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <W12>  
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone F25H8  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19949  
A:Accession: T24896  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <W12>  
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone T13H10  
C:Genetics:  
A:Gene: CESP:F25H8.3  
A:Map position: 4  
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81

Query Match 4.9%; Score 124; DB 2; Length 2165;  
Best Local Similarity 20.0%; Pred. No. 0.33; Mismatches 129; Indels 164; Gaps 21;  
Matches 83; Conservative 39;  
Qy 141 KDHPENKPSVSPDWR-----AWQRLSLARANSQDQYXDSTSD-- 186  
Db 1177 KPCHRESCPKYL--GEMSQCSVCEGDWSSRRVSCVSGNGTEVDMSLCGTASDRPASHQ 1234  
Qy 187 -SNFLNPP--RGWDHTA-----PGHRTFTKQDPEYDSTDGE--GD----- 222  
Db 1235 TCNLGTCPFFWRNTDWSACSVCGIGHRETTCTIYREQSDASFCGDTKMPETSQTCHLL 1294  
Qy 223 ----W--SLASVCSVTCGNGNKRTRSC--GYACTATRESRTCDRPNCPGIEDTFTTAATE 274  
Db 1295 PCTSWKFSHSPSCVTCGSGIGTQHSVSCTRGSEGTIVDEYFCFDRNTRPLRKKT----- 1347  
Qy 275 VSLLAGSEEFNATKLFVEVDYDTSCE-----RWMSCKS-----E 306  
Db 1348 -----CERDTCGPRVLQKLOADVPIRWATGPWTACATCNGTGQR 1390  
Qy 307 FLKYMVKVMNDLPSCPCSPYPTVAYSTADIFDIKRKDFRWKDAAGPKLEIYKPTAR 366  
Db 1391 LLKCRDH--VRDLPDEXCNH-----LDKEVSTRNCLRDCS-----YMKMAE 1430

Qy 367 YCIRSMLSLESTLLAAHCYCGDNMQLITRGAGCTPNLISFEFAEL-----HYKV 418  
Db 1431 W-----EECPATCGTHVQOSRNVTCVSAEDGGRT---ILKDVCDDVQKRPTSARNCEL 1480  
Qy 419 DVLP-----WII-----CKGWSR-----YNEARPPNNGOKCTE 447  
Db 1481 EPCPKGEBHIGWIIGDWSKCSASC GGWRRRSVCTSSSCDETRKPKMFKCNE 1535  
RESULT 10  
T00355  
hypothetical protein KIAA0688 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00355  
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl A:Reference number: Z14142; MUID:98403880  
A:Accession: T00355  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-837 <I5H>  
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190  
A:Experimental source: brain  
C:Genetics:  
A:Gene: KIAA0688  
C:Superfamily: thrombospondin type 1 repeat homology  
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>  
Query Match 4.8%; Score 122; DB 2; Length 837;  
Best Local Similarity 22.3%; Pred. No. 0.14;  
Matches 61; Conservative 21; Mismatches 88; Indels 104; Gaps 11;  
Qy 104 LLDLPNFPDLKADINGOXPNIQVTIEVVDGPDSEADKQDQ-----PENKPSMSVSPDWR 159  
Db 424 LLDKPEAP-----LHLPVTF---PGKDYDADQCOLTFGDSRHCPQLPPCAA 469  
Qy 160 AMQRLSLARANSQDQYXDSTSDSNFLNPPRGWDHTAPGHRTFTTKDQPEYDST-- 217  
Db 470 LWC-----SGHLN-----GHAMCQTKHSPWADGTPC 495  
Qy 218 -----DGEQWSLWSVCTCGNGNKRTRSC----- 244  
Db 496 GPAQACMGRCRLHMDQLQDFNIPQAGWPGWPGWDCSRTCGGVQVFSRDCRTRPVPRNG 555  
Qy 245 GYACTA--TESRTCDRPNCP--GIEDTFTTAATEVSLLAGSEEFNATKLFVEVDYDTSCE-- 300  
Db 556 GKYPEGRTFRSCNTEDCPTGSALTFR-----EEQCAAYNHRTDLFKSFPQPMWVPRYT 611  
Qy 301 -----MSCKSEFLKKYMHKVMNDLPSCPCS 325  
Db 612 GVAPDQCKLTQCARALGYVYVLEPRVVVDGTPCS 645  
RESULT 11  
A42587  
thrombospondin 2 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Aug-1999  
C:Accession: A42587; A39851  
R:Lahter, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M. J. Biol. Chem. 267, 3274-3281, 1992  
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce A:Reference number: A42587; MUID:92147683  
A:Accession: A42587  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1172 <LAH>  
A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241  
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)

[illegible]



submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid F08C6.  
A:Reference number: Z18440

A:Accession: U15976

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-957 <BEN>

A:Cross-references: EMBL:U29378; NID:g868184; PID:g868185; PIDN:AAA68721.1; CESP:F08C6.1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F08C6.1

A:Introns: 23/1; 135/3; 220/3; 285/3; 325/2; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7

Query Match 4.7%; Score 118; DB 2; Length 957;

Best Local Similarity 23.3%; Pred. No. 0.35;

Matches 58; Conservative 27; Mismatches 82; Indels 82; Gaps 14;

Qy 221 GDWSLWSCVTCGNGNOKRTRSCGYA--CTATESRT--CDRPNCGIEDTFTRTAATEVS 276

Db 608 GTWSLWTSCTATCGGGYKRNKACISITGCEGNEDETEVCSSSECSVS----- 655

Qy 277 LLAGE-----EFNATKLFVDTDSC-----ERWMSCKS-----EFLKKYMHKVMNDLPS 321

Db 656 LRVGNWSTWTEWNCV-----SCGRSQARYKCLSPHRTLAFDCPGENKVTNEL-- 707

Qy 322 CPCSYPTVAVSTADIF-----DRIKKDFRWDA-----SGPKKLEIYKPTARYC 368

Db 708 -----RITFFKARSYMCVSRCKNKRNTISEKNIEVRSCDNGPCNAIGVGTGWGWS 760

Qy 369 IRSMLSLESTTLAQHCYGDNMQLITRGKAGTGNLSTFSAELHYKVDYLPWLICKG 428

Db 761 TCS-TSCGPGTLVRQTC-----NREPCDGSAHERRSCNV---ATCQN 799

Qy 429 D--WSRYNE 435

Db 800 DGIWSLWNE 808

RESULT 15

A38152

F-spondin - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Nov-2000

C:Accession: A38152

R:Klar A.; Baldassare, M.; Jessell, T. M.

Cell 69, 95-110, 1992

A:Title: F-spondin: a gene expressed at high levels in the floor plate encodes a secreted

A:Reference number: A38152; MUID:92208952

A:Accession: A38152

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-807 <KLA>

A:Cross-references: GB:M88469; NID:g204176; PIDN:AAA41174.1; PID:g204177

A:Experimental source: embryo floor plate

A:Note: sequence extracted from NCBI backbone (NCBIN:90877, NCBI:90878)

C:Superfamily: F-spondin: thrombospondin type 1 repeat homology

F:441-495/Domain: thrombospondin type 1 repeat homology <THR2>

F:500-555/Domain: thrombospondin type 1 repeat homology <THR3>

F:557-611/Domain: thrombospondin type 1 repeat homology <THR4>

F:613-666/Domain: thrombospondin type 1 repeat homology <THR1>

F:667-721/Domain: thrombospondin type 1 repeat homology <THR5>

F:753-807/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 4.7%; Score 117.5; DB 2; Length 807;

Best Local Similarity 21.0%; Pred. No. 0.31;

Matches 82; Conservative 35; Mismatches 137; Indels 137; Gaps 17;

Qy 83 RPRFRQETGHPSLQRDFRSLFLDLPNFPD-----LSKADINGQXPNI---QVT 128

Db 368 KPTIPQEKIRPLTSLDHQSPFYD-----PEGGSITQVARVVIERTARKQECQCNIVPDNDV 423

Qy 129 IEVVDGPDSEADKQDHP-----NKPSW-----SVFSPD- 157

Db 424 DIVADLAPPEKEDDDTPTFCIYSNWSWWSACSSSTCEKGRMRQRLKQALDLSVPCPDT 483

Qy 158 -----WRAWQWRSLSLARANSQDQDYXYDSTSDSDSNFLNPPR 194

Db 484 QDFQPCMGPGCDEDEGSTCTMSEWITWSPCSVSCGNGMRSRERYVKQFFPEDGVCMLP-- 541

Qy 195 GWDHTAPGHRFTETKDOPEYDSTDGEG---DWLSLWSVCSVTCGNGNOKRTR-----SC 244

Db 542 -----TEETKCTVNECSPSSCLVTEWGEWDDCSATCGMGKMKRRHRVVKMSPAD 591

Qy 245 GYACTA--TESRTCDRPNCPGIEDTFTAATEVSLIAGSEEFNATKLFVEDTDCSERWMS 302

Db 592 GSMCKAETSQAEEKMMPECHTIP-----CLLSPWSEWSDCSV-----TCGKGMR 635

Qy 303 CKSEFLK-----KYMHKVMNDLPSCP--CSYPTVAYNSTAD-----IFDR 340

Db 636 TRQRLKSLAEILGDCNEDLEQAEKCH--LPECPIDCELSEWSQWSECNKSCKGKGMIRTR 693

Qy 341 IKRKDFRWKDSGPKKLEIYKPTARYCIRS 371

Db 694 TIQMEPQFGGACPCP-ETVORKKCRARKCLRS 723

Search completed: April 25, 2002, 17:20:19

Job time: 12041 sec





KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
KW SIGNAL. 1 20  
FT CHAIN 21 1572  
FT DOMAIN 21 924  
FT TRANSMEM 925 945  
FT DOMAIN 946 953  
FT TRANSMEM 954 974  
FT DOMAIN 975 982  
FT TRANSMEM 983 1003  
FT DOMAIN 1004 1024  
FT TRANSMEM 1025 1045  
FT DOMAIN 1046 1066  
FT TRANSMEM 1067 1087  
FT DOMAIN 1088 1141  
FT TRANSMEM 1142 1162  
FT DOMAIN 1163 1168  
FT TRANSMEM 1169 1189  
FT DOMAIN 1190 1572  
FT DOMAIN 297 351  
FT DOMAIN 352 406  
FT DOMAIN 407 461  
FT DOMAIN 463 517  
FT DOMAIN 859 911  
FT DOMAIN 117 122  
FT DOMAIN 177 180  
FT DOMAIN 222 225  
FT DOMAIN 1303 1306  
FT DOMAIN 1352 1358  
FT DOMAIN 1413 1418  
FT CARBOHYD 94 94  
FT CARBOHYD 179 179  
FT CARBOHYD 180 180  
FT CARBOHYD 344 344  
FT CARBOHYD 425 425  
FT CARBOHYD 548 548  
FT CARBOHYD 633 633  
FT CARBOHYD 855 855  
SQ SEQUENCE 1572 AA; 171140 MW; A9775645B7BC285 CRC64;  
  
Query Match 5.7%; Score 143; DB 1; Length 1572;  
Best Local Similarity 19.5%; Pred. No. 0.0028;  
Matches 89; Conservative 47; Mismatches 161; Indels 160; Gaps 19;  
  
QY 114 SKADINQXPNIQVTIEVVDGPDSEADKQHPENKPSW--SVSP-----Dw 158  
DB 248 AEADLHSGSSNDLFTTMRVGEPE---EPPKVKTPWPSADEPGLYNAQTGDPAAEW 303  
QY 159 RAWQBSLSLARNSGDQYXDSTSDSNFL-----NPPRGWDHTAPGHTFE---- 207  
DB 304 SPMSVCSLT---CGGGLQVTRTSCVSPSYGTLCGSLRTPCNNSATCPVHGVEWGS 360  
QY 208 -----TKDQPEYDSTDGEG-----DWSLWSVCS 230  
DB 361 WSLCSRSGRSGRSRMRTCPVPOHGKACGEBELQYKLCMACPVEGOWLEWGPWGPCS 420  
QY 231 VTGCGNQKRTSCGYA-----CTA--TESRTCDPNCPIGIEDTERTAAVEVSLLAGSE 282  
DB 421 TSCANTQQRKCSVAGPAWATCTGALDTRCSNLECPATDSKW-----G 467  
QY 283 EFNATKFE--VDTDCSRWMSCKSEFLKYHKYHKVNDLPSCPCSYPTVEAYSTAD---- 336  
DB 468 PWNWSLCSKTCDTGQWRFRMCQATGTQGY-----PCEGTGEVVKPCSEKRC 516  
QY 337 IPDRIKDFE---RKWDASGPEKELEIKPTA-----RYCIRS----- 371  
DB 517 AFHEMCRDEYVLMVTKKAAAGEIYKNCPPNAGSASRCLLSAQGVAYWGLPSPFARCI 576  
QY 372 -----MLSLESTTLAAQHCYCGDNQMLITRGKAGTGNLIS--TEFSAELHYKVDVLP 422  
DB 577 SHEYRYLYLSLREHLAKGQMLAGEGMSQVVR-----SLQELLARTYYSGLDLFSVDIL- 631

QY 423 WIICKDWSRYNEARPPNNGQKCTESPSDEDIYKQFQ 459  
DB 632 -----RNVTDTKRATYVPSADDDVQRFQ 655  
  
RESULT 2  
ID BAIL\_HUMAN STANDARD; PRT: 1584 AA.  
AC O14514;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Brain-specific angiogenesis inhibitor 1 precursor.  
GN BAIL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Fetal brain;  
RX MEDLINE=98054121; PubMed=9393972;  
RA Nishimori H., Shiratsuchi T., Urano T., Kimura Y., Kiyono K.,  
RA Tatsumi K., Yoshida S., Ono M., Kuwano M., Nakamura Y., Tokino T.;  
RT "A novel brain-specific p53-target gene, BAIL, containing  
Thrombospondin type 1 repeats inhibits experimental angiogenesis.";  
RL Oncogene 15:2145-2150(1997).  
RN [2]  
RP INTERACTION WITH BAP1.  
RX MEDLINE=98321173; PubMed=9647739;  
RA Shiratsuchi T., Futamura M., Oda K., Nishimori H., Nakamura Y.,  
RA Tokino T.;  
RT "Cloning and characterization of BAI-associated protein 1: a PDZ  
domain-containing protein that interacts with BAIL.";  
RL Biochem. Biophys. Res. Commun. 247:597-604(1998).  
CC -!- FUNCTION: LIKELY TO BE A POTENT INHIBITOR OF ANGIOGENESIS IN  
BRAIN AND MAY PLAY A SIGNIFICANT ROLE AS A MEDIATOR OF THE P53  
SIGNAL IN SUPPRESSION OF GLOBLASTOMA. MAY FUNCTION IN CELL  
ADHESION AND SIGNAL TRANSDUCTION IN THE BRAIN.  
CC -!- SUBUNIT: INTERACTS WITH BAP1.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LIKELY TO BE  
CONCENTRATED AT CELL-CELL ADHESION SITES.  
CC -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN. REDUCED OR NO  
EXPRESSION IS OBSERVED IN SOME GLIOBLASTOMA CELL LINES AND CANCER  
TISSUES.  
CC -!- INDUCTION: BY P53.  
CC -!- DOMAIN: THE TSP1 REPEATS INHIBIT IN VIVO ANGIOGENESIS IN RAT  
CORNEA INDUCED BY BFGF.  
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AB005297; BAA23647.1; -  
DR MIM; 602682; -  
DR InterPro; IPR000832; GPCR\_secretin.  
DR InterPro; IPR000203; PKD\_cys\_rich.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR001879; hormn\_receptor.  
DR Pfam; PF00002; 7tm\_2; 1.  
DR Pfam; PF01825; GPS; 1.  
DR Pfam; PF02793; HRM; 1.  
DR Pfam; PF00090; tsp\_1; 5.  
DR SMART; SM00303; GPS; 1.  
DR SMART; SM00008; Hormr; 1.  
DR SMART; SM00209; TSP1; 5.  
DR

DR PROSITE; PS50221; GPS; 1.  
DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2.1; FALSE\_NEG.  
DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2.2; FALSE\_NEG.  
DR PROSITE; PS50227; G\_PROTEIN\_RECEP\_F2.3; 1.  
DR PROSITE; PS50261; G\_PROTEIN\_RECEP\_F2.4; 1.  
DR PROSITE; PS50092; TSP1; 5.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
KW Repeat; Cell adhesion.  
FT SIGNAL 1 30  
FT CHAIN 31 1584  
FT DOMAIN 31 948  
FT TRANSMEM 949 969  
FT DOMAIN 970 980  
FT TRANSMEM 981 1001  
FT DOMAIN 1002 1008  
FT TRANSMEM 1009 1029  
FT DOMAIN 1030 1052  
FT TRANSMEM 1053 1073  
FT DOMAIN 1074 1093  
FT TRANSMEM 1094 1114  
FT DOMAIN 1115 1136  
FT TRANSMEM 1137 1157  
FT DOMAIN 1158 1166  
FT TRANSMEM 1167 1187  
FT DOMAIN 1188 1198  
FT TRANSMEM 1199 1219  
FT DOMAIN 1220 1240  
FT TRANSMEM 1241 1261  
FT DOMAIN 1262 1282  
FT TRANSMEM 1283 1303  
FT DOMAIN 1304 1324  
FT TRANSMEM 1325 1345  
FT DOMAIN 1346 1366  
FT TRANSMEM 1367 1387  
FT DOMAIN 1388 1408  
FT TRANSMEM 1409 1429  
FT DOMAIN 1430 1450  
FT TRANSMEM 1451 1471  
FT DOMAIN 1472 1492  
FT TRANSMEM 1493 1513  
FT DOMAIN 1514 1534  
FT TRANSMEM 1535 1555  
FT DOMAIN 1556 1576  
FT TRANSMEM 1577 1597  
FT DOMAIN 1598 1618  
FT TRANSMEM 1619 1639  
FT DOMAIN 1640 1660  
FT TRANSMEM 1661 1681  
FT DOMAIN 1682 1702  
FT TRANSMEM 1703 1723  
FT DOMAIN 1724 1744  
FT TRANSMEM 1745 1765  
FT DOMAIN 1766 1786  
FT TRANSMEM 1787 1807  
FT DOMAIN 1808 1828  
FT TRANSMEM 1829 1849  
FT DOMAIN 1850 1870  
FT TRANSMEM 1871 1891  
FT DOMAIN 1892 1912  
FT TRANSMEM 1913 1933  
FT DOMAIN 1934 1954  
FT TRANSMEM 1955 1975  
FT DOMAIN 1976 1996  
FT TRANSMEM 1997 2017  
FT DOMAIN 2018 2038  
FT TRANSMEM 2039 2059  
FT DOMAIN 2060 2080  
FT TRANSMEM 2081 2101  
FT DOMAIN 2102 2122  
FT TRANSMEM 2123 2143  
FT DOMAIN 2144 2164  
FT TRANSMEM 2165 2185  
FT DOMAIN 2186 2206  
FT TRANSMEM 2207 2227  
FT DOMAIN 2228 2248  
FT TRANSMEM 2249 2269  
FT DOMAIN 2270 2290  
FT TRANSMEM 2291 2311  
FT DOMAIN 2312 2332  
FT TRANSMEM 2333 2353  
FT DOMAIN 2354 2374  
FT TRANSMEM 2375 2395  
FT DOMAIN 2396 2416  
FT TRANSMEM 2417 2437  
FT DOMAIN 2438 2458  
FT TRANSMEM 2459 2479  
FT DOMAIN 2480 2500  
FT TRANSMEM 2501 2521  
FT DOMAIN 2522 2542  
FT TRANSMEM 2543 2563  
FT DOMAIN 2564 2584  
FT TRANSMEM 2585 2605  
FT DOMAIN 2606 2626  
FT TRANSMEM 2627 2647  
FT DOMAIN 2648 2668  
FT TRANSMEM 2669 2689  
FT DOMAIN 2690 2710  
FT TRANSMEM 2711 2731  
FT DOMAIN 2732 2752  
FT TRANSMEM 2753 2773  
FT DOMAIN 2774 2794  
FT TRANSMEM 2795 2815  
FT DOMAIN 2816 2836  
FT TRANSMEM 2837 2857  
FT DOMAIN 2858 2878  
FT TRANSMEM 2879 2899  
FT DOMAIN 2900 2920  
FT TRANSMEM 2921 2941  
FT DOMAIN 2942 2962  
FT TRANSMEM 2963 2983  
FT DOMAIN 2984 3004  
FT TRANSMEM 3005 3025  
FT DOMAIN 3026 3046  
FT TRANSMEM 3047 3067  
FT DOMAIN 3068 3088  
FT TRANSMEM 3089 3109  
FT DOMAIN 3110 3130  
FT TRANSMEM 3131 3151  
FT DOMAIN 3152 3172  
FT TRANSMEM 3173 3193  
FT DOMAIN 3194 3214  
FT TRANSMEM 3215 3235  
FT DOMAIN 3236 3256  
FT TRANSMEM 3257 3277  
FT DOMAIN 3278 3298  
FT TRANSMEM 3299 3319  
FT DOMAIN 3320 3340  
FT TRANSMEM 3341 3361  
FT DOMAIN 3362 3382  
FT TRANSMEM 3383 3403  
FT DOMAIN 3404 3424  
FT TRANSMEM 3425 3445  
FT DOMAIN 3446 3466  
FT TRANSMEM 3467 3487  
FT DOMAIN 3488 3508  
FT TRANSMEM 3509 3529  
FT DOMAIN 3530 3550  
FT TRANSMEM 3551 3571  
FT DOMAIN 3572 3592  
FT TRANSMEM 3593 3613  
FT DOMAIN 3614 3634  
FT TRANSMEM 3635 3655  
FT DOMAIN 3656 3676  
FT TRANSMEM 3677 3697  
FT DOMAIN 3698 3718  
FT TRANSMEM 3719 3739  
FT DOMAIN 3740 3760  
FT TRANSMEM 3761 3781  
FT DOMAIN 3782 3802  
FT TRANSMEM 3803 3823  
FT DOMAIN 3824 3844  
FT TRANSMEM 3845 3865  
FT DOMAIN 3866 3886  
FT TRANSMEM 3887 3907  
FT DOMAIN 3908 3928  
FT TRANSMEM 3929 3949  
FT DOMAIN 3950 3970  
FT TRANSMEM 3971 3991  
FT DOMAIN 3992 4012  
FT TRANSMEM 4013 4033  
FT DOMAIN 4034 4054  
FT TRANSMEM 4055 4075  
FT DOMAIN 4076 4096  
FT TRANSMEM 4097 4117  
FT DOMAIN 4118 4138  
FT TRANSMEM 4139 4159  
FT DOMAIN 4160 4180  
FT TRANSMEM 4181 4201  
FT DOMAIN 4202 4222  
FT TRANSMEM 4223 4243  
FT DOMAIN 4244 4264  
FT TRANSMEM 4265 4285  
FT DOMAIN 4286 4306  
FT TRANSMEM 4307 4327  
FT DOMAIN 4328 4348  
FT TRANSMEM 4349 4369  
FT DOMAIN 4370 4390  
FT TRANSMEM 4391 4411  
FT DOMAIN 4412 4432  
FT TRANSMEM 4433 4453  
FT DOMAIN 4454 4474  
FT TRANSMEM 4475 4495  
FT DOMAIN 4496 4516  
FT TRANSMEM 4517 4537  
FT DOMAIN 4538 4558  
FT TRANSMEM 4559 4579  
FT DOMAIN 4580 4600  
FT TRANSMEM 4601 4621  
FT DOMAIN 4622 4642  
FT TRANSMEM 4643 4663  
FT DOMAIN 4664 4684  
FT TRANSMEM 4685 4705  
FT DOMAIN 4706 4726  
FT TRANSMEM 4727 4747  
FT DOMAIN 4748 4768  
FT TRANSMEM 4769 4789  
FT DOMAIN 4790 4810  
FT TRANSMEM 4811 4831  
FT DOMAIN 4832 4852  
FT TRANSMEM 4853 4873  
FT DOMAIN 4874 4894  
FT TRANSMEM 4895 4915  
FT DOMAIN 4916 4936  
FT TRANSMEM 4937 4957  
FT DOMAIN 4958 4978  
FT TRANSMEM 4979 4999  
FT DOMAIN 5000 5020  
FT TRANSMEM 5021 5041  
FT DOMAIN 5042 5062  
FT TRANSMEM 5063 5083  
FT DOMAIN 5084 5104  
FT TRANSMEM 5105 5125  
FT DOMAIN 5126 5146  
FT TRANSMEM 5147 5167  
FT DOMAIN 5168 5188  
FT TRANSMEM 5189 5209  
FT DOMAIN 5210 5230  
FT TRANSMEM 5231 5251  
FT DOMAIN 5252 5272  
FT TRANSMEM 5273 5293  
FT DOMAIN 5294 5314  
FT TRANSMEM 5315 5335  
FT DOMAIN 5336 5356  
FT TRANSMEM 5357 5377  
FT DOMAIN 5378 5398  
FT TRANSMEM 5399 5419  
FT DOMAIN 5420 5440  
FT TRANSMEM 5441 5461  
FT DOMAIN 5462 5482  
FT TRANSMEM 5483 5503  
FT DOMAIN 5504 5524  
FT TRANSMEM 5525 5545  
FT DOMAIN 5546 5566  
FT TRANSMEM 5567 5587  
FT DOMAIN 5588 5608  
FT TRANSMEM 5609 5629  
FT DOMAIN 5630 5650  
FT TRANSMEM 5651 5671  
FT DOMAIN 5672 5692  
FT TRANSMEM 5693 5713  
FT DOMAIN 5714 5734  
FT TRANSMEM 5735 5755  
FT DOMAIN 5756 5776  
FT TRANSMEM 5777 5797  
FT DOMAIN 5798 5818  
FT TRANSMEM 5819 5839  
FT DOMAIN 5840 5860  
FT TRANSMEM 5861 5881  
FT DOMAIN 5882 5902  
FT TRANSMEM 5903 5923  
FT DOMAIN 5924 5944  
FT TRANSMEM 5945 5965  
FT DOMAIN 5966 5986  
FT TRANSMEM 5987 6007  
FT DOMAIN 6008 6028  
FT TRANSMEM 6029 6049  
FT DOMAIN 6050 6070  
FT TRANSMEM 6071 6091  
FT DOMAIN 6092 6112  
FT TRANSMEM 6113 6133  
FT DOMAIN 6134 6154  
FT TRANSMEM 6155 6175  
FT DOMAIN 6176 6196  
FT TRANSMEM 6197 6217  
FT DOMAIN 6218 6238  
FT TRANSMEM 6239 6259  
FT DOMAIN 6260 6280  
FT TRANSMEM 6281 6301  
FT DOMAIN 6302 6322  
FT TRANSMEM 6323 6343  
FT DOMAIN 6344 6364  
FT TRANSMEM 6365 6385  
FT DOMAIN 6386 6406  
FT TRANSMEM 6407 6427  
FT DOMAIN 6428 6448  
FT TRANSMEM 6449 6469  
FT DOMAIN 6470 6490  
FT TRANSMEM 6491 6511  
FT DOMAIN 6512 6532  
FT TRANSMEM 6533 6553  
FT DOMAIN 6554 6574  
FT TRANSMEM 6575 6595  
FT DOMAIN 6596 6616  
FT TRANSMEM 6617 6637  
FT DOMAIN 6638 6658  
FT TRANSMEM 6659 6679  
FT DOMAIN 6680 6700  
FT TRANSMEM 6701 6721  
FT DOMAIN 6722 6742  
FT TRANSMEM 6743 6763  
FT DOMAIN 6764 6784  
FT TRANSMEM 6785 6805  
FT DOMAIN 6806 6826  
FT TRANSMEM 6827 6847  
FT DOMAIN 6848 6868  
FT TRANSMEM 6869 6889  
FT DOMAIN 6890 6910  
FT TRANSMEM 6911 6931  
FT DOMAIN 6932 6952  
FT TRANSMEM 6953 6973  
FT DOMAIN 6974 6994  
FT TRANSMEM 6995 7015  
FT DOMAIN 7016 7036  
FT TRANSMEM 7037 7057  
FT DOMAIN 7058 7078  
FT TRANSMEM 7079 7099  
FT DOMAIN 7100 7120  
FT TRANSMEM 7121 7141  
FT DOMAIN 7142 7162  
FT TRANSMEM 7163 7183  
FT DOMAIN 7184 7204  
FT TRANSMEM 7205 7225  
FT DOMAIN 7226 7246  
FT TRANSMEM 7247 7267  
FT DOMAIN 7268 7288  
FT TRANSMEM 7289 7309  
FT DOMAIN 7310 7330  
FT TRANSMEM 7331 7351  
FT DOMAIN 7352 7372  
FT TRANSMEM 7373 7393  
FT DOMAIN 7394 7414  
FT TRANSMEM 7415 7435  
FT DOMAIN 7436 7456  
FT TRANSMEM 7457 7477  
FT DOMAIN 7478 7498  
FT TRANSMEM 7499 7519  
FT DOMAIN 7520 7540  
FT TRANSMEM 7541 7561  
FT DOMAIN 7562 7582  
FT TRANSMEM 7583 7603  
FT DOMAIN 7604 7624  
FT TRANSMEM 7625 7645  
FT DOMAIN 7646 7666  
FT TRANSMEM 7667 7687  
FT DOMAIN 7688 7708  
FT TRANSMEM 7709 7729  
FT DOMAIN 7730 7750  
FT TRANSMEM 7751 7771  
FT DOMAIN 7772 7792  
FT TRANSMEM 7793 7813  
FT DOMAIN 7814 7834  
FT TRANSMEM 7835 7855  
FT DOMAIN 7856 7876  
FT TRANSMEM 7877 7897  
FT DOMAIN 7898 7918  
FT TRANSMEM 7919 7939  
FT DOMAIN 7940 7960  
FT TRANSMEM 7961 7981  
FT DOMAIN 7982 8002  
FT TRANSMEM 8003 8023  
FT DOMAIN 8024 8044  
FT TRANSMEM 8045 8065  
FT DOMAIN 8066 8086  
FT TRANSMEM 8087 8107  
FT DOMAIN 8108 8128  
FT TRANSMEM 8129 8149  
FT DOMAIN 8150 8170  
FT TRANSMEM 8171 8191  
FT DOMAIN 8192 8212  
FT TRANSMEM 8213 8233  
FT DOMAIN 8234 8254  
FT TRANSMEM 8255 8275  
FT DOMAIN 8276 8296  
FT TRANSMEM 8297 8317  
FT DOMAIN 8318 8338  
FT TRANSMEM 8339 8359  
FT DOMAIN 8360 8380  
FT TRANSMEM 8381 8401  
FT DOMAIN 8402 8422  
FT TRANSMEM 8423 8443  
FT DOMAIN 8444 8464  
FT TRANSMEM 8465 8485  
FT DOMAIN 8486 8506  
FT TRANSMEM 8507 8527  
FT DOMAIN 8528 8548  
FT TRANSMEM 8549 8569  
FT DOMAIN 8570 8590  
FT TRANSMEM 8591 8611  
FT DOMAIN 8612 8632  
FT TRANSMEM 8633 8653  
FT DOMAIN 8654 8674  
FT TRANSMEM 8675 8695  
FT DOMAIN 8696 8716  
FT TRANSMEM 8717 8737  
FT DOMAIN 8738 8758  
FT TRANSMEM 8759 8779  
FT DOMAIN 8780 8800  
FT TRANSMEM 8801 8821  
FT DOMAIN 8822 8842  
FT TRANSMEM 8843 8863  
FT DOMAIN 8864 8884  
FT TRANSMEM 8885 8905  
FT DOMAIN 8906 8926  
FT TRANSMEM 8927 8947  
FT DOMAIN 8948 8968  
FT TRANSMEM 8969 8989  
FT DOMAIN 8990 9010  
FT TRANSMEM 9011 9031  
FT DOMAIN 9032 9052  
FT TRANSMEM 9053 9073  
FT DOMAIN 9074 9094  
FT TRANSMEM 9095 9115  
FT DOMAIN 9116 9136  
FT TRANSMEM 9137 9157  
FT DOMAIN 9158 9178  
FT TRANSMEM 9179 9199  
FT DOMAIN 9200 9220  
FT TRANSMEM 9221 9241  
FT DOMAIN 9242 9262  
FT TRANSMEM 9263 9283  
FT DOMAIN 9284 9304  
FT TRANSMEM 9305 9325  
FT DOMAIN 9326 9346  
FT TRANSMEM 9347 9367  
FT DOMAIN 9368 9388  
FT TRANSMEM 9389 9409  
FT DOMAIN 9410 9430  
FT TRANSMEM 9431 9451  
FT DOMAIN 9452 9472  
FT TRANSMEM 9473 9493  
FT DOMAIN 9494 9514  
FT TRANSMEM 9515 9535  
FT DOMAIN 9536 9556  
FT TRANSMEM 9557 9577  
FT DOMAIN 9578 9598  
FT TRANSMEM 9599 9619  
FT DOMAIN 9620 9640  
FT TRANSMEM 9641 9661  
FT DOMAIN 9662 9682  
FT TRANSMEM 9683 9703  
FT DOMAIN 9704 9724  
FT TRANSMEM 9725 9745  
FT DOMAIN 9746 9766  
FT TRANSMEM 9767 9787  
FT DOMAIN 9788 9808  
FT TRANSMEM 9809 9829  
FT DOMAIN 9830 9850  
FT TRANSMEM 9851 9871  
FT DOMAIN 9872 9892  
FT TRANSMEM 9893 9913  
FT DOMAIN 9914 9934  
FT TRANSMEM 9935 9955  
FT DOMAIN 9956 9976  
FT TRANSMEM 9977 9997  
FT DOMAIN 10000 10020  
FT TRANSMEM 10021 10041  
FT DOMAIN 10042 10062  
FT TRANSMEM 10063 10083  
FT DOMAIN 10084 10104  
FT TRANSMEM 10105 10125  
FT DOMAIN 10126 10146  
FT TRANSMEM 10147 10167  
FT DOMAIN 10168 10188  
FT TRANSMEM 10189 10209  
FT DOMAIN 10210 10230  
FT TRANSMEM 10231 10251  
FT DOMAIN 10252 10272  
FT TRANSMEM 10273 10293  
FT DOMAIN 10294 10314  
FT TRANSMEM 10315 10335  
FT DOMAIN 10336 10356  
FT TRANSMEM 10357 10377  
FT DOMAIN 10378 10398  
FT TRANSMEM 10399 10419  
FT DOMAIN 10420 10440  
FT TRANSMEM 10441 10461  
FT DOMAIN 10462 10482  
FT TRANSMEM 10483 10503  
FT DOMAIN 10504 10524  
FT TRANSMEM 10525 10545  
FT DOMAIN 10546 10566  
FT TRANSMEM 10567 10587  
FT DOMAIN 10588 10608  
FT TRANSMEM 10609 10629  
FT DOMAIN 10630 10650  
FT TRANSMEM 10651 10671  
FT DOMAIN 10672 10692  
FT TRANSMEM 10693 10713  
FT DOMAIN 10714 10734  
FT TRANSMEM 10735 10755  
FT DOMAIN 10756 10776  
FT TRANSMEM 10777 10797  
FT DOMAIN 10798 10818  
FT TRANSMEM 10819 10839  
FT DOMAIN 10840 10860  
FT TRANSMEM 10861 10881  
FT DOMAIN 10882 10902  
FT TRANSMEM 10903 10923  
FT DOMAIN 10924 10944  
FT TRANSMEM 10945 10965  
FT DOMAIN 10966 10986  
FT TRANSMEM 10987 11007  
FT DOMAIN 11008 11028  
FT TRANSMEM 11029 11049  
FT DOMAIN 11050 11070  
FT TRANSMEM 11071 11091  
FT DOMAIN 11092 11112  
FT TRANSMEM 11113 11133  
FT DOMAIN 11134 11154  
FT TRANSMEM 11155 11175  
FT DOMAIN 11176 11196  
FT TRANSMEM 11197 11217  
FT DOMAIN 11218 11238  
FT TRANSMEM 11239 11259  
FT DOMAIN 11260 11280  
FT TRANSMEM 11281 11301  
FT DOMAIN 11302 11322  
FT TRANSMEM 11323 11343  
FT DOMAIN 11344 11364  
FT TRANSMEM 11365 11385  
FT DOMAIN 11386 11406  
FT TRANSMEM 11407 11427  
FT DOMAIN 11428 11448  
FT TRANSMEM 11449 11469  
FT DOMAIN 11470 11490  
FT TRANSMEM 11491 11511  
FT DOMAIN 11512 11532  
FT TRANSMEM 11533 11553  
FT DOMAIN 11554 11574  
FT TRANSMEM 11575 11595  
FT DOMAIN 11596 11616  
FT TRANSMEM 11617 11637  
FT DOMAIN 11638 11658  
FT TRANSMEM 11659 11679  
FT DOMAIN 11680 11700  
FT TRANSMEM 11701 11721  
FT DOMAIN 11722 11742  
FT TRANSMEM 11743 11763  
FT DOMAIN 11764 11784  
FT TRANSMEM 11785 11805  
FT DOMAIN 11806 11826  
FT TRANSMEM 11827 11847  
FT DOMAIN 11848 11868  
FT TRANSMEM 11869 11889  
FT DOMAIN 11890 11910  
FT TRANSMEM 11911 11931  
FT DOMAIN 11932 11952  
FT TRANSMEM 11953 11973  
FT DOMAIN 11974 11994  
FT TRANSMEM 11995 12015  
FT DOMAIN 12016 12036  
FT TRANSMEM 12037 12057  
FT DOMAIN 12058 12078  
FT TRANSMEM 12079 12099  
FT DOMAIN 12100 12120  
FT TRANSMEM 12121 12141  
FT DOMAIN 12142 12162  
FT TRANSMEM 12163 12183  
FT DOMAIN 12184 12204  
FT TRANSMEM 12205 12225  
FT DOMAIN 12226 12246  
FT TRANSMEM 12247 12267  
FT DOMAIN 12268 12288  
FT TRANSMEM 12289 12309  
FT DOMAIN 12310 12330  
FT TRANSMEM 12331 12351  
FT DOMAIN 12352 12372  
FT TRANSMEM 12373 12393  
FT DOMAIN 12394 12414  
FT TRANSMEM 12415 12435  
FT DOMAIN 12436 12456  
FT TRANSMEM 12457 12477  
FT DOMAIN 12478 12498  
FT TRANSMEM 12499 12519  
FT DOMAIN 12520 12540  
FT TRANSMEM 12541 12561  
FT DOMAIN 12562 12582  
FT TRANSMEM 12583 12603  
FT DOMAIN 12604 12624  
FT TRANSMEM 12625 12645  
FT DOMAIN 12646 12666  
FT TRANSMEM 12667 12687  
FT DOMAIN 12688 12708  
FT TRANSMEM 12709 12729  
FT DOMAIN 12730 12750  
FT TRANSMEM 12751 12771  
FT DOMAIN 12772 12792  
FT TRANSMEM 12793 12813  
FT DOMAIN 12814 12834  
FT TRANSMEM 12835 12855  
FT DOMAIN 12856 12876  
FT TRANSMEM 12877 12897  
FT DOMAIN 12898 12918  
FT TRANSMEM 12919 12939  
FT DOMAIN 12940 12960  
FT TRANSMEM 12961 12981  
FT DOMAIN 12982 13002  
FT TRANSMEM 13003 13023  
FT DOMAIN 13024 13044  
FT TRANSMEM 13045 13065  
FT DOMAIN 13066 13086  
FT TRANSMEM 13087 13107  
FT DOMAIN 13108 13128  
FT TRANSMEM 13129 13149  
FT DOMAIN 13150 13170  
FT TRANSMEM 13171 13191  
FT DOMAIN 13192 13212  
FT TRANSMEM 13213 13233  
FT DOMAIN 13234 13254  
FT TRANSMEM 13255 13275  
FT DOMAIN 13276 13296  
FT TRANSMEM 13297 13317  
FT DOMAIN 13318 13338  
FT TRANSMEM 13339 13359  
FT DOMAIN 13360 13380  
FT TRANSMEM 13381 13401  
FT DOMAIN 13402 13422  
FT TRANSMEM 13423 13443  
FT DOMAIN 13444 13464  
FT TRANSMEM 13465 13485  
FT DOMAIN 13486 13506  
FT TRANSMEM 13507 13527  
FT DOMAIN 13528 13548  
FT TRANSMEM 13549 13569  
FT DOMAIN 13570 13590  
FT TRANSMEM 13591 13611  
FT DOMAIN 13612 13632  
FT TRANSMEM 13633 13653  
FT DOMAIN 13654 13674  
FT TRANSMEM 13675 13695  
FT DOMAIN 13696 13716  
FT TRANSMEM 13717 13737  
FT DOMAIN 13738 13758  
FT TRANSMEM 13759 13779  
FT DOMAIN 13780 13800  
FT TRANSMEM 13801 13821  
FT DOMAIN 13822 13842  
FT TRANSMEM 13843 13863  
FT DOMAIN 13864 13884  
FT TRANSMEM 13885 13905  
FT DOMAIN 13906 13926  
FT TRANSMEM 13927 13947  
FT DOMAIN 13948 13968  
FT TRANSMEM 13969 13989  
FT DOMAIN 13990 14010  
FT TRANSMEM 14011 14031  
FT DOMAIN 14032 14052  
FT TRANSMEM 14053 14073  
FT DOMAIN 14074 14094  
FT TRANSMEM 14095 14114  
FT DOMAIN 14115 14135  
FT TRANSMEM 14136 14156  
FT DOMAIN 14157 14177  
FT TRANSMEM 14178 14198  
FT DOMAIN 14199 14219  
FT TRANSMEM 14220 14240  
FT DOMAIN 14241 14261  
FT TRANSMEM 14262 14282  
FT DOMAIN 14283 14303  
FT TRANSMEM 14304 14324  
FT DOMAIN 14325 14345  
FT TRANSMEM 14346 14366  
FT DOMAIN 14367 14387  
FT TRANSMEM 14388 14408  
FT DOMAIN 14409 14429  
FT TRANSMEM 14430 14450  
FT DOMAIN 14451 14471  
FT TRANSMEM 14472 14492  
FT DOMAIN 14493 14513  
FT TRANSMEM 14514 14534  
FT DOMAIN 14535 14555  
FT TRANSMEM 14556 14576  
FT DOMAIN 14577 14597  
FT TRANSMEM 14598 14618  
FT DOMAIN 14619 14639  
FT TRANSMEM 14640 14660  
FT DOMAIN 14661 14681  
FT TRANSMEM 14682 14702  
FT DOMAIN 14703 14723  
FT TRANSMEM 14724 14744  
FT DOMAIN 14745 14765  
FT TRANSMEM 14766 14786  
FT DOMAIN 14787 14807  
FT TRANSMEM 14808 14828  
FT DOMAIN 14829 14849  
FT TRANSMEM 14850 14870  
FT DOMAIN 14871 14891  
FT TRANSMEM 14892 14912  
FT DOMAIN 14913 14933  
FT TRANSMEM 14934 14954  
FT DOMAIN 14955 14975  
FT TRANSMEM 14976 14996  
FT DOMAIN 14997 15017  
FT DOMAIN 15018 15038  
FT TRANSMEM 15039 15059  
FT DOMAIN 15060 15080  
FT TRANSMEM 15081 15101  
FT DOMAIN 15102 15122  
FT TRANSMEM 15123 15143  
FT DOMAIN 15144 15164  
FT TRANSMEM 15165 15185  
FT DOMAIN 15186 15206  
FT TRANSMEM 15207 15227  
FT DOMAIN 15228 15248  
FT TRANSMEM 15249 15269  
FT DOMAIN 15270 15290  
FT TRANSMEM 15291 15311  
FT DOMAIN 15312 15332  
FT TRANSMEM 15333 15353  
FT DOMAIN 15354 15374  
FT TRANSMEM 15375 15395  
FT DOMAIN 15396 15416  
FT TRANSMEM 15417 15437  
FT DOMAIN 15438 15458  
FT TRANSMEM 15459 15479  
FT DOMAIN 15480 15500  
FT TRANSMEM 15501 15521  
FT DOMAIN 15522 15542  
FT TRANSMEM 15543 15563  
FT DOMAIN 15564 15584  
FT TRANSMEM 15585 15605  
FT DOMAIN 15606 15626  
FT TRANSMEM 15627 15647  
FT DOMAIN 15648 15668  
FT TRANSMEM 15669 15689  
FT DOMAIN 15690 15710  
FT TRANSMEM 15711 15731  
FT DOMAIN 15732 15752  
FT TRANSMEM 15753 15773  
FT DOMAIN 15774 15794  
FT TRANSMEM 15795 15815  
FT DOMAIN 15816 15836  
FT TRANSMEM 15837 15857  
FT DOMAIN 15858 15878  
FT TRANSMEM 15879 15899  
FT DOMAIN 15900 15920  
FT TRANSMEM 15921 15941  
FT DOMAIN 15942 15962  
FT TRANSMEM 15963 15983  
FT DOMAIN 15984 16004  
FT TRANSMEM 16005 16025  
FT DOMAIN 16026 16046  
FT TRANSMEM 16047 16067  
FT DOMAIN 16068 16088  
FT TRANSMEM 16089 16109  
FT DOMAIN 16110 16130  
FT TRANSMEM 16131 16151  
FT DOMAIN 16152 16172  
FT TRANSMEM 16173 16193  
FT DOMAIN 16194 16214  
FT TRANSMEM 16215 16235  
FT DOMAIN 16236 16256  
FT TRANSMEM 16257 16277  
FT DOMAIN 16278 16298  
FT TRANSMEM 16299 16319  
FT DOMAIN 16320 16340  
FT TRANSMEM 16341 16361  
FT DOMAIN 16362 16382  
FT TRANSMEM 16383 16403  
FT DOMAIN 16404 16424  
FT TRANSMEM 16425 16445  
FT DOMAIN 16446 16466  
FT TRANSMEM 16467 16487  
FT DOMAIN 16488 16508  
FT TRANSMEM 16509 16529  
FT DOMAIN 16530 16550  
FT TRANSMEM 16551 16571  
FT DOMAIN 16572 16592  
FT TRANSMEM 16593 16613  
FT DOMAIN 16614 16634  
FT TRANSMEM 16635 16655  
FT DOMAIN 16656 16676  
FT TRANSMEM 16677 16697  
FT DOMAIN 16698 16718  
FT TRANSMEM 16719 16739  
FT DOMAIN 16740 16760  
FT TRANSMEM 16761 16781  
FT DOMAIN 16782 16802  
FT TRANSMEM 16803 16823  
FT DOMAIN 1682

DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50092; TSP1; 3.  
 DR PROSITE; PS01208; VWC; 1.  
 DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
 KW EGF-like domain; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 1173  
 FT DOMAIN 23 335  
 FT DOMAIN 319 376  
 FT DOMAIN 382 433  
 FT DOMAIN 438 494  
 FT DOMAIN 456 546  
 FT DOMAIN 550 590  
 FT DOMAIN 591 648  
 FT DOMAIN 649 693  
 FT DOMAIN 726 761  
 FT DOMAIN 762 784  
 FT DOMAIN 785 820  
 FT DOMAIN 821 843  
 FT DOMAIN 844 881  
 FT DOMAIN 882 917  
 FT DOMAIN 918 953  
 FT DOMAIN 954 1173  
 FT SITE 929 931  
 FT DISULFID 554 565  
 FT DISULFID 559 575  
 FT DISULFID 578 589  
 FT DISULFID 595 611  
 FT DISULFID 602 620  
 FT DISULFID 623 647  
 FT DISULFID 653 666  
 FT DISULFID 660 679  
 FT DISULFID 681 692  
 FT CARBOHYD 155 158  
 FT CARBOHYD 158 158  
 FT CARBOHYD 250 250  
 FT CARBOHYD 363 363  
 FT CARBOHYD 705 705  
 FT CARBOHYD 711 711  
 FT CARBOHYD 1070 1070  
 FT SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64;  
 Query Match 5.4%; Score 135; DB 1; Length 1173;  
 Best Local Similarity 22.3%; Pred. No. 0.0083;  
 Matches 59; Conservative 23; Mismatches 88; Indels 94; Gaps 13;  
 Qy 135 PDSEADKQHPENKPSWSPDW-----RAWMQSRSLARANGDQDYXDSTS-- 184  
 Db 377 PSDSADDDWSP-----WS-----DWTGCVTCGHGQGRSCDNLNPPCGSSVQTRSCQ 427  
 Qy 185 --DDSNFLNPRGWDHTAPGHRFTET-----KDQEYDSTDGEGD----- 222  
 Db 428 IQDCDKRFQDQGWHSWSPSSCVTCGSGQITRILCNLSPVQLNGKCEGEGRENKPC 487  
 Qy 223 -----WSLWVSCVTCGNGNQKTRSC-----GYACTA--TESRTCDRP 259  
 Db 488 QKDCPINGQWPSLMDTCTVTCGGQMKRRLCNLPKQYEGKDCIGETDSQICNKQ 547  
 Qy 260 NCP-----GIEDTFTAAATEVSLLAGSEE-----FNATKLFVDVTDSCRRW 301  
 Db 548 DCPIDGCLSNPCFAGVRCV-----SFIDGSKWKCSCPPGYRGNGITCKDIDECKEVP 599  
 Qy 302 -SKSEFLKMYHVMNDLSPCC 324  
 Db 600 DAC---FTLNGVHRCENTEPGYN 620  
 RESULT 4  
 ID\_BAI3\_HUMAN STANDARD; -PRT; 1522 AA.

AC 060242; 060297;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Brain-specific angiogenesis inhibitor 3 precursor.  
 GN BAI3 OR KIAA0550.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM).  
 RP TISSUE=Fetal brain;  
 RC MEDLINE=98194217; PubMed=9533023;  
 RA Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;  
 RT "Cloning and characterization of BAI2 and BAI3, novel genes homologous  
 RT to brain-specific angiogenesis inhibitor 1 (BAI1).";  
 RL Cytogenet. Cell Genet. 79:103-108(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RP TISSUE=Brain;  
 RC MEDLINE=98290545; PubMed=9628581;  
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,  
 RA Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RL code for large proteins in vitro.";  
 RL DNA Res. 5:31-39(1998).  
 CC -!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION AND  
 CC SUPPRESSION OF GLOBLASTOMA.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN  
 CC HEART. REDUCED EXPRESSION IS OBSERVED IN SOME GLOBLASTOMA CELL  
 CC LINES.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB005299; BAA25363.1; -;  
 DR EMBL; AB011122; BAA25476.1; -;  
 DR MIM; 602684; -;  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR000203; PKD\_cys\_rich.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00090; tsp\_1; 4.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00008; Hormr; 1.  
 DR SMART; SM00209; TSP1; 4.  
 DR PROSITE; PS01180; CUB; 1.  
 DR PROSITE; PS02221; GPS; 1.  
 DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
 DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
 DR PROSITE; PS02227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS02261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE; PS00092; TSP1; 4.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
 KW Repeat; Alternative splicing.

```
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1522 BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 3.
FT DOMAIN 25 880 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 881 901 1 (POTENTIAL).
FT DOMAIN 902 910 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 911 931 2 (POTENTIAL).
FT DOMAIN 932 939 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 940 960 3 (POTENTIAL).
FT DOMAIN 961 981 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 982 1002 4 (POTENTIAL).
FT DOMAIN 1003 1023 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1024 1044 5 (POTENTIAL).
FT DOMAIN 1045 1098 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1099 1119 6 (POTENTIAL).
FT DOMAIN 1120 1125 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1126 1146 7 (POTENTIAL).
FT DOMAIN 1147 1522 CYTOPLASMIC (POTENTIAL).
FT CUB. 30 159
FT DOMAIN 291 344 TSP TYPE-1 1.
FT TRANSMEM 345 399 TSP TYPE-1 2.
FT DOMAIN 400 454 TSP TYPE-1 3.
FT TRANSMEM 455 509 TSP TYPE-1 4.
FT DOMAIN 816 868 GRS..
FT TRANSMEM 942 945 POLY-THR.
FT DOMAIN 1173 1176
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 82 82 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 812 812 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 828 828 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 937 937 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 643 665 MISSING (IN SHORT ISOFORM).
FT VARSPLIC 990 1007 LPAVVATSVGTRTGY -> KHIDIPHALKMLNTH
                               (IN SHORT ISOFORM).
FT VARSPLIC 1008 1522 MISSING (IN SHORT ISOFORM).
FT SEQUENCE 1522 AA; 171490 MW; D22D0A5D4BB62502 CRC64;
```

Query Match 5.3%; Score 134.5; DB 1; Length 1522;  
Best Local Similarity 21.4%; Pred. No. 0.013;  
Matches 84; Conservative 48; Mismatches 170; Indels 91; Gaps 21;

```
QY 62 SLSEA---PREHL---DHQAAHQPFRFRQETGHPSLQDFRPSFLD-----LPN 109
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 NLTREARPKPEFGMGMDHTIKS---RPRSVHEKRVPOEQADAAR-FNAQTGESGVEE 296
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 FPLSKADIN-GQXPNQVITIEV-----VDGPDSEADKDQHPENKPSWSVPSPDRAWW 162
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 297 WSQWSTCSVTGCGSOVRTRTCVSPYTHCSGPLRESRVNNTALCPVHGV---WEEWS 352
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 163 QRLSLARNSGDQDYXSDTSDDSNFLNPPRGWDHTAPGHRFTFKDQPEYSDTGE-G 221
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 353 PWSLCSFTCGRGORTTRSC-----PPQYGGPRCEGPETHHKPCNTALCPVDGQWQ 404
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 222 DWSLWSVSVTCGNGNKRTRSC-----GYACTA---TESRTCDRPNCPGIEDTFRTAAT 273
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 405 EWSWSQCSVTCNGTGOORSQCTAAAHGSGEGRGPAWESRECVNPEC----- 452
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 274 EVSLLAGSEEFNATKLFVDTDSC-----RWMSCKSEFLK-KYMHKVMNDLPSCP---C 324
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 453 -----TANGQWNGHWSGCKSCDGGWERRIRTCQCAVITGQCEGTGEVRRCSQRQ 507
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 325 SYTEVA---YSTADIFDRKKRDFRWK----DASGPKK---LEIY-----KPTARYC 368
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 508 PAPYEICPEDYLSMWNKRTPAAGDLAFNQCPLNATGTTSRRCSLSLGHVAFWEQPSPARC 567
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 369 IRS-----MLSLESTTLAAQHCYCCYGDNMQLITR 396
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 568 ISNEYRHLQHSIKHLAKGQMLAGCMGMSQVTK 600
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
PROP_MOUSE
ID PROP_MOUSE STANDARD; PRT; 437 AA.
AC P11680;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Properdin (Fragment).
OS PFC.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Macrophage;
RX MEDLINE=88318954; PubMed=3045564;
RA Goundis A., Reid K.B.M.;
RT "Properdin, the terminal complement components, thrombospondin and
   the circumsporozoite protein of malaria parasites contain similar
   sequence motifs.";
RL Nature 335:82-85(1988).
CC -1- FUNCTION: A POSITIVE REGULATOR OF THE ALTERNATE PATHWAY OF
   COMPLEMENT. IT BINDS TO AND STABILIZES THE C3-AND C5-CONVERTASE
   ENZYME COMPLEXES.
CC -1- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X12905; CAA31389.1; -.
CC PIR; S05478; S05478.
CC MGD; MGI:97545; PFC.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF000090; tsp.1; 6.
CC SMART; SM00209; TSP1; 5.
CC PROSITE; PS00092; TSP1; 6.
CC Complement alternate pathway; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN 46 103 TSP TYPE-1 1.
FT DOMAIN 104 160 TSP TYPE-1 2.
FT DOMAIN 161 224 TSP TYPE-1 3.
FT DOMAIN 225 282 TSP TYPE-1 4.
FT DOMAIN 283 345 TSP TYPE-1 5.
FT DOMAIN 346 405 TSP TYPE-1 6.
FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 437 AA; 47538 MW; 2B8DBCE22B3B78BE CRC64;
```

Query Match 5.2%; Score 131; DB 1; Length 437;  
Best Local Similarity 26.5%; Pred. No. 0.005;  
Matches 53; Conservative 16; Mismatches 79; Indels 52; Gaps 11;

```
QY 155 SPDWRWQWSLSLARNSGDQDYXSDTSDDSNFLNPPRGW---DHTAPGHRTFE---T 208
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 46 SPQWSAWSLWGPCSVTCSEGSQLRHRCVG-----RGQCSSEVAPGTLQWQLOAC 96
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 209 KDQPEYSDTDEGWSLWSVCSVTCGNGNKRTR-----SCGYAC--TATESRTCD-R 258
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 97 EDQPCCEMGWSWGWGPGCVTCSCGTQIRQVCONPAPKCGCHPGCAQSQACDTQ 156
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 259 PNCPIGIEDTFRTAAT-----EVSLLAGSEEFNATKLFVDYDTSCE-----WMSC 303
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 157 KTCP-----THGAWASMGWPSRSGSLGGAQBPKEKTR-----SRSCAPAPSHQPPGKPC 207

QY 304 KSEFLAKYHKYHNDLPSCP 323

Db 208 SG----PAYEHKACSGLPSCP 224

RESULT 6

AT52\_HUMAN

ID AT52\_HUMAN STANDARD; PRT; 1211 AA.

AC O95450;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and

DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)

DE (procollagen 1/II amino-propeptide processing enzyme) (Procollagen I

DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)

DE (procollagen 1/II amino-propeptide processing enzyme).

GN ADAMTS2 OR PCINP OR PCPNI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI).

RC TISSUE=Skin;

RC MEDLINE=99347993; PubMed=10417273;

RA Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,

RA Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,

RA Byers P.H., Lapierre C.M., Prockop D.J., Nussgens B.V.;

RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis

RT are caused by mutations in the procollagen I N-proteinase gene.";

RL Am. J. Hum. Genet. 65:308-317(1999).

CC -I- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR

CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO

CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDENT OF ITS ROLE IN

CC COLLAGEN BIOSYNTHESIS.

CC -I- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain

CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains

CC at Ala-1-Gln.

CC -I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -I- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO

CC COLLAGEN TYPE XIV (BY SIMILARITY).

CC -I- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR

CC MATRIX (BY SIMILARITY).

CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;

CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-

CC PROCOLLAGEN PEPTIDASE ACTIVITY.

CC -I- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON

CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

CC -I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT

CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -I- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY

CC SIMILARITY).

CC -I- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS

CC SYNDROME TYPE VIIC (EDS-VIIC), A RECESSIVELY INHERITED DISORDER

CC CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND

CC BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY

CC PROCESSED AT THE AMINO TERMINUS.

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -I- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -I- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.

CC -I- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

DR EMBL; AJ003125; CAA05880.1; --

DR MEROPS; M12.301; --

DR MIM; 604539; --

DR MIM; 225410; --

DR InterPro; IPR001762; Disintegrin.

DR InterPro; IPR002870; Pep\_M12B\_propep.

DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSP1.

DR InterPro; IPR000130; Zn\_Mtpeptidse.

DR Pfam; PF01562; Pep\_M12B\_propep; 1.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF00090; Tsp1; 4.

DR SMART; SM00209; TSP1; 4.

DR PROSITE; PS50215; ADAM\_MEPRO; 1.

DR PROSITE; PS50092; TSP1; 1.

DR PROSITE; PS00142; ZINC\_PROTEASE; FALSE\_NEG.

DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;

KW Alternative splicing.

FT SIGNAL 1 29

FT PROPEP 30 253

FT CHAIN 254 1211

FT METAL 408 408

FT ACT\_SITE 409 409

FT METAL 412 412

FT METAL 418 418

FT DOMAIN 480 560

FT DOMAIN 561 617

FT DOMAIN 618 722

FT SITE 691 693

FT DOMAIN 723 851

FT DOMAIN 852 911

FT DOMAIN 912 974

FT DOMAIN 975 1030

FT DOMAIN 40 43

FT DOMAIN 185 188

FT CARBOHYD 112 112

FT CARBOHYD 251 251

FT CARBOHYD 949 949

FT CARBOHYD 993 993

FT CARBOHYD 1031 1031

FT CARBOHYD 1098 1098

FT CARBOHYD 1145 1145

FT CARBOHYD 1150 1150

FT VARSPIC 544 566

FT VARSPIC 567 1211

FT SEQUENCE 1211 AA; 134722 MW; BECEE25C23CAD2D CRC64;

QY 192 PPRGWDHTAPGHTFTKDK---OPEYDSTDGE-GDMSLWCVSVTCGNGKQKTRSC---244

Db 532 PPLDGTMCAPGKHCFCGHCILWLTPDLILKRDGSGWANGSPFGSCSRTCTGTVKFTQCDNP 591

QY 245 -----GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVDTDC 297

Db 592 HPANGGRTCSGLAYDFQLCSRDQCPDLSADFR-----EEQC 627

QY 298 ERWMSCKSEFLKMYHMKVMDLSPCSCSYTEVAYSTADIFDKIKRDFRW-----KDas 352

Db 628 RQW-----DLYFEHGDAAQHHLWLPHEIRDA- 651

QY 353 GPKREKLEIYKPTARYCIRSMLSLESTTLAAHCYCGDNMOLLITRG--KGAGTLENLSTEF 410

Db 652 --KERCHLYCESRE--TGEVYSMKRMVHDTRCSYKDAFSLCVGRGCKRVGCGVIG---704

QY 411 SAEILHYKVDVLPWLICKGWSRYNEARPPNNGOKCTESPSEDIYIKOFQ 459

Query Match 5.2%; Score 130; DB 1; Length 1211;

Best Local Similarity 21.8%; Pred. No. 0.021;

Matches 63; Conservative 26; Mismatches 102; Indels 98; Gaps 13;



```

Db 705 SSKQEDKCGV-----CGGDNHCKVVKV-----GTFTSPKKHGYIKMFE 743
RESULT 7
TSP2_BOVIN
ID TSP2_BOVIN STANDARD; PRT: 1170 AA.
AC Q95116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor (Corticotropin-induced secreted protein)
DE (CISP).
GN THBS2 OR TSP2 OR TSP-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
RP SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lafeuillade M., Keramidas M., Aguesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 1-522 FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells.";
RL J. Cell. Physiol. 167:164-172(1996).
[3]
RP SEQUENCE OF 318-831 FROM N.A.
RC TISSUE=ortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta.";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 WFCC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96540; CAA65385.1; -.
DR EMBL; X87620; CAA60952.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WFCC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.

```

---

```

DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; WFCC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
EGF-like domain; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1170 THROMBOSPONDIN 2.
FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 318 375 WFCC.
FT DOMAIN 379 430 TSP TYPE-1 1.
FT DOMAIN 435 491 TSP TYPE-1 2.
FT DOMAIN 492 546 TSP TYPE-1 3.
FT DOMAIN 547 587 EGF-LIKE 1.
FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 646 690 EGF-LIKE 3.
FT DOMAIN 723 758 TSP TYPE-3 1.
FT DOMAIN 759 781 TSP TYPE-3 2.
FT DOMAIN 782 817 TSP TYPE-3 3.
FT DOMAIN 818 840 TSP TYPE-3 4.
FT DOMAIN 841 878 TSP TYPE-3 5.
FT DOMAIN 879 914 TSP TYPE-3 6.
FT DOMAIN 915 950 TSP TYPE-3 7.
FT DOMAIN 951 1170 C-TERMINAL.
FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 551 562 BY SIMILARITY.
FT DISULFID 556 572 BY SIMILARITY.
FT DISULFID 575 586 BY SIMILARITY.
FT DISULFID 592 608 BY SIMILARITY.
FT DISULFID 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT DISULFID 657 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 535 535 A -> V (IN REF. 3).
FT CONFLICT 748 748 S -> T (IN REF. 3).
SQ SEQUENCE 1170 AA; 129862 MW; 9CF1FBF5B89A051 CRC64;

```

---

```

Query Match 5.1%; Score 128.5; DB 1; Length 1170;
Best Local Similarity 23.1%; Pred. No. 0.027;
Matches 61; Conservative 18; Mismatches 86; Indels 99; Gaps 14;

QY 146 ENKPSWSPVPDWRWQRSLARANSQDQYXYDSTSDSNFLNP----- 192
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 DGEEGHS----PWAETECs---ATCGSGTQGRSCDVTSTNCLGPSIOTRACSLGRCD 429
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 193 -----PRGWDHTAP-----GHRFTF-----ETKD-QPEYD 215
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 HRIRODGGWHSWSPSSCVTCGVGNVTRILNCNSPVQPMQGRSKSGSRETRACQGPC 489
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 STDGE-GDWSLSWCVSTCGNGNQKTRSC-----GYACT--ATESRTCDRPNC--- 261
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 PVDGRWSPWSPSACTVTTCAGGIRIRVCNPSPEPHGGKDCYGGAKEQOMCRKSCPID 549
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 -----PGIE-DTERTAAEVSLLAGSEFNATKLFVYD-----TDCERWMSCKSEF 307
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 GCLSNPCFFGAECSSFPDGSWSCGCPGGFLGNTHCEDLDECAVVDVC-----F 600
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 308 LKKYMHKVMNDLPS-----CPCSY 326
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
Db 601 ATSKAHCYNTNPGYHCLPCPPRY 624
RESULT 8
TSPL_BOVIN
ID TSPL_BOVIN STANDARD; PRT; 1170 AA.
AC Q28178; Q28179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSP1 OR TSP-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactylia; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HOLSTEIN; TISSUE-Tooth;
RX MEDLINE=98173773; PubMed=9507054;
RA Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
RA Inoue H.;
RT "cDNA cloning of bovine thrombospondin 1 and its expression in
RT odontoblasts and predentin.";
RL Biochim. Biophys. Acta 1382:17-22(1998).
RN [2]
RP SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
RC TISSUE=Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC V/BETA-3 AND ALPHA-1IB/BETA-3. MAY PLAY A ROLE IN DENTINOGENESIS
CC AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- TISSUE SPECIFICITY: ODONTOBLASTS.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB005287; BAA21115.1; -.
DR EMBL; X87618; CAA60950.1; -.
DR EMBL; X87619; CAA60951.1; -.
DR HSSP; P35555; LEMO.
DR GlycoSuiteDB; Q28178; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 16.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.

DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS01208; VWFC; 1.
DR PROSITE; PS00992; TSP1; 3.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232
FT DOMAIN 316 373
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 548
FT DOMAIN 549 587
FT DOMAIN 588 645
FT DOMAIN 646 690
FT DOMAIN 733 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 270 270
FT DISULFID 274 274
FT DISULFID 551 562
FT DISULFID 556 572
FT DISULFID 575 586
FT DISULFID 592 608
FT DISULFID 599 617
FT DISULFID 620 644
FT DISULFID 650 663
FT DISULFID 657 676
FT DISULFID 678 689
FT DISULFID 688 689
FT CARBOHYD 248 248
FT CARBOHYD 360 360
FT CARBOHYD 708 708
FT CARBOHYD 1067 1067
FT CARBOHYD 1085 1085
FT CONFLICT 805 805
FT SEQUENCE 1170 AA; 129533 MW; 0DD6ADF3E5FA031A CRC64;

Query Match 5.1%; Score 127.5; DB 1; Length 1170;
Best Local Similarity 25.6%; Pred. No. 0.032;
Matches 45; Conservative 10; Mismatches 56; Indels 65; Gaps 9;

QY 151 WSPVSPD--WRW--W-----QRLSLARAN-----SGQDXYXSDSDSNF 189
DB 373 WPSDSADGGSPWSEWTSVTCGNGIQGRSCDILNNRCEGSSVQTRTCHIQECDKRF 432
QY 190 LNPFRGMDHTAPGH-----RTFETK--DQPEYDS 216
DB 433 -KODGWSHSPWSSCSVTCGGVITRILCNLSPSPQMNGKPCGEGRKATKCKOSCP 491
QY 217 TDGEGDWSLWSVCVTCGNGNQKRTSC-----GYACT--ATESRTCDRNP 262
DB 492 NGWGPWSPWDICSVTCGGVGQKRSRLCNPKPQFGGKDCVGDVTENQICNKQDCP 547

RESULT 9
TSPL_MOUSE
ID TSPL_MOUSE STANDARD; PRT; 1170 AA.
AC P35441;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92128941; PubMed=1774063;  
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,  
 RA Jenkins N.A.;  
 RT "Characterization of the murine thrombospondin gene.";  
 RL Genomics 11:587-600(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92147683; PubMed=1371115;  
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,  
 RA Dixit V.M.;  
 RT "Characterization of mouse thrombospondin 2 sequence and expression  
 during cell growth and development.";  
 RL J. Biol. Chem. 267:3274-3281(1992).  
 RN [3]  
 RP SEQUENCE OF 1-490 FROM N.A.  
 RX MEDLINE=90375346; PubMed=2398070;  
 RA Bornstein P., Alfi D., Devarayalu S., Framson P., Li P.;  
 RT "Characterization of the mouse thrombospondin gene and evaluation of  
 the role of the first intron in human gene expression.";  
 RL J. Biol. Chem. 265:16691-16698(1990).  
 CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-  
 V/BETA-3 AND ALPHA-1IB/BETA-3.  
 CC -!- SUBUNIT: HOMOPRIMER; DISULFIDE-LINKED.  
 CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 WFCC DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M62450; AAA50611.1; JOINED.  
 DR EMBL; M62451; AAA50611.1; JOINED.  
 DR EMBL; M62452; AAA50611.1; JOINED.  
 DR EMBL; M62453; AAA50611.1; JOINED.  
 DR EMBL; M62454; AAA50611.1; JOINED.  
 DR EMBL; M62455; AAA50611.1; JOINED.  
 DR EMBL; M62456; AAA50611.1; JOINED.  
 DR EMBL; M62457; AAA50611.1; JOINED.  
 DR EMBL; M62458; AAA50611.1; JOINED.  
 DR EMBL; M62459; AAA50611.1; JOINED.  
 DR EMBL; M62460; AAA50611.1; JOINED.  
 DR EMBL; M62461; AAA50611.1; JOINED.  
 DR EMBL; M62462; AAA50611.1; JOINED.  
 DR EMBL; M62463; AAA50611.1; JOINED.  
 DR EMBL; M62464; AAA50611.1; JOINED.  
 DR EMBL; M62465; AAA50611.1; JOINED.  
 DR EMBL; M62466; AAA50611.1; JOINED.  
 DR EMBL; M62467; AAA50611.1; JOINED.  
 DR EMBL; M62468; AAA50611.1; JOINED.  
 DR EMBL; M62469; AAA50611.1; JOINED.  
 DR EMBL; M87276; AAA53063.1; --  
 DR EMBL; J05606; AAA40431.1; --  
 DR EMBL; J05605; AAA40431.1; JOINED.  
 DR PIR; A40558; A40558.  
 DR PIR; B42587; B42587.  
 DR HSSP; P35555; 1EMN.  
 DR MGD; MGI:98737; Thbs1.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF\_Ca.

DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR001007; WFCC.  
 DR InterPro; IPR003367; tsp\_3.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF00090; tsp\_1; 3.  
 DR Pfam; PF02412; tsp\_3; 8.  
 DR Pfam; PF00093; wvc; 1.  
 DR SMART; SM00181; EGF; 3.  
 DR SMART; SM00209; TSP1; 3.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00092; TSP1; 3.  
 DR PROSITE; PS01208; WFCC; 1.  
 DR KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
 EGF-like domain; Signal.  
 FT SIGNAL 1 18 POTENTIAL  
 FT CHAIN 19 1170 THROMBOSPONDIN 1  
 FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).  
 FT DOMAIN 316 373 WFCC.  
 FT DOMAIN 379 430 TSP TYPE-1 1.  
 FT DOMAIN 435 491 TSP TYPE-1 2.  
 FT DOMAIN 492 548 TSP TYPE-1 3.  
 FT DOMAIN 549 587 EGF-LIKE 1.  
 FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 646 690 EGF-LIKE 3.  
 FT DOMAIN 723 758 TSP TYPE-3 1.  
 FT DOMAIN 759 781 TSP TYPE-3 2.  
 FT DOMAIN 782 817 TSP TYPE-3 3.  
 FT DOMAIN 818 840 TSP TYPE-3 4.  
 FT DOMAIN 841 878 TSP TYPE-3 5.  
 FT DOMAIN 879 914 TSP TYPE-3 6.  
 FT DOMAIN 915 950 TSP TYPE-3 7.  
 FT DOMAIN 951 1170 C-TERMINAL.  
 FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DISULFID 270 270 INTERCHAIN (PROBABLE).  
 FT DISULFID 274 274 INTERCHAIN (PROBABLE).  
 FT DISULFID 551 562 BY SIMILARITY.  
 FT DISULFID 556 572 BY SIMILARITY.  
 FT DISULFID 575 586 BY SIMILARITY.  
 FT DISULFID 592 608 BY SIMILARITY.  
 FT DISULFID 599 617 BY SIMILARITY.  
 FT DISULFID 620 644 BY SIMILARITY.  
 FT DISULFID 650 663 BY SIMILARITY.  
 FT DISULFID 657 676 BY SIMILARITY.  
 FT DISULFID 678 689 BY SIMILARITY.  
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 1025 1025 F -> L (IN REF. 2).  
 SQ SEQUENCE 1170 AA; 129646 MW; 0443E493615E7F06 CRC64;

Query Match 5.1%; Score 127.5; DB 1; Length 1170;  
 Best Local Similarity 23.3%; Pred. No. 0.032;  
 Matches 41; Conservative 17; Mismatches 53; Indels 65; Gaps 8;  
 QY 151 WSVPSPD--WRAWWORSLSLARANSQDQ--YXYVSTSD-----DSNF 189  
 Db 373 WPSDSADDGHSWSPSEWTSCTGCGIQQGRSCDSLNNRCSSSVQVTRTCHIQECDKRF 432  
 QY 190 LNPFRGWDHTAPGHRFTET-----KDOPEYDSTDEGD----- 222  
 Db 433 -KQDGGWSHSPWSSCSVTCGDGVITRILCNSPSPOMNGKPCGEARETKACKKDACP 491  
 QY 223 -----WSLWSVCSVTCGNGNQKRTSC-----GYACT--ATESRTCDRNCP 262  
 Db 492 NGGWPSPWDICSVTCGGVGVRRLCNNTPTQFGGKDCVGDVTENQVCNKQDCP 547

```

RESULT 10
TSPL_HUMAN
ID TSPL_HUMAN STANDARD; PRT; 1170 AA.
AC P07996;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSPL OR TSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Endothelial cells;
RA Lawler J., Hynes R.O.;
RX MEDLINE=87057617; PubMed=2430973;
RT "The structure of human thrombospondin, an adhesive glycoprotein with
RT multiple calcium-binding sites and homologies with several different
RT proteins.";
RL J. Cell Biol. 103:1635-1648(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89139590; PubMed=2918029;
RA Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,
RA Baumgartel D.M., Rotwein P., Frazier W.A.;
RT "Complete thrombospondin mRNA sequence includes potential regulatory
RT sites in the 3' untranslated region.";
RL J. Cell Biol. 108:729-736(1989).
RN [3]
RP SEQUENCE OF 1-397 FROM N.A.
RX MEDLINE=87157592; PubMed=3030396;
RA Kobayashi S., Eden-McCuchan F., Framson P., Bornstein P.;
RT "Partial amino acid sequence of human thrombospondin as determined by
RT analysis of cDNA clones: homology to malarial circumsporozoite
RT proteins.";
RL Biochemistry 25:8418-8425(1986).
RN [4]
RP SEQUENCE OF 1-374 FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [5]
RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=89291870; PubMed=2544587;
RA Laherty C.D., Gierman T.M., Dixit V.M.;
RT "Characterization of the promoter region of the human thrombospondin
RT gene. DNA sequences within the first intron increase transcription.";
RL J. Biol. Chem. 264:11222-11227(1989).
RN [6]
RP SEQUENCE OF 1028-1170 FROM N.A.
RA la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
RT Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CARBOHYDRATE-LINKAGE SITES W-385; S-394; W-438; W-441; T-450; W-498
AND T-507.
RC TISSUE-Platelet;
RX PubMed=11067851;
RA Hofsteenge J., Huwiler K.G., Macek B., Hess D., Lawler J.,
RA Mosher D.F., Peter-Katalinic J.;
RT "C-mannosylation and O-fucosylation of the thrombospondin type 1
RT module.";
RL J. Biol. Chem. 276:6485-6498(2001).
CC -1- FUNCTION: Adhesive glycoprotein that mediates cell-to-cell and
CC cell-to-matrix interactions. Can bind to fibrinogen, fibronectin,
CC laminin, type V collagen and integrins alpha-V/beta-1, alpha-
CC V/beta-3 and alpha-IIIb/beta-3.
CC -1- SUBUNIT: Homotrimer; disulfide-linked.

```

```

CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25631; AAA36741.1; -
DR EMBL; X04665; CAA28370.1; -
DR EMBL; X14787; CAA32889.1; -
DR EMBL; J04835; AAA61178.1; -
DR EMBL; M99425; AAB59366.1; -
DR PIR; A05172; A05172.
DR PIR; A25812; A25812.
DR PIR; A26155; A26155.
DR PIR; A30140; A30140.
DR PIR; A34274; A34274.
DR HSP; P35555; IEMN.
DR GlycoSuiteDB; P07996; -
DR MIM; 188060; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; VWFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232.
FT DOMAIN 316 373.
FT DOMAIN 379 430.
FT DOMAIN 435 491.
FT DOMAIN 492 548.
FT DOMAIN 549 587.
FT DOMAIN 588 645.
FT DOMAIN 646 690.
FT DOMAIN 723 758.
FT DOMAIN 759 781.
FT DOMAIN 782 817.
FT DOMAIN 818 840.
FT DOMAIN 841 878.
FT DOMAIN 879 914.
FT DOMAIN 915 950.
FT DOMAIN 951 1170.
FT SITE 926 928.
FT DISULFID 270 270.
FT DISULFID 274 274.
FT DISULFID 551 562.
FT DISULFID 556 572.
FT DISULFID 575 586.

```

THROMBOSPONDIN 1.  
HEPARIN-BINDING (POTENTIAL).

VWFC.

TSP TYPE-1 1.

TSP TYPE-1 2.

TSP TYPE-1 3.

EGF-LIKE 1.

EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

EGF-LIKE 3.

TSP TYPE-3 1.

TSP TYPE-3 2.

TSP TYPE-3 3.

TSP TYPE-3 4.

TSP TYPE-3 5.

TSP TYPE-3 6.

TSP TYPE-3 7.

C-TERMINAL.

CELL ATTACHMENT SITE (POTENTIAL).

INTERCHAIN (PROBABLE).

INTERCHAIN (PROBABLE).

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

```

FT DISULFID 592 608 BY SIMILARITY.
FT FT 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT FT 644 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT FT 663 663 BY SIMILARITY.
FT DISULFID 657 676 BY SIMILARITY.
FT FT 676 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT FT 689 689 BY SIMILARITY.
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .).
FT FT 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 C-LINKED (MAN).
FT FT 360 360 C-LINKED (MAN).
FT CARBOHYD 385 385 /FTid=CAR_000205.
FT FT 385 385 O-LINKED (FUC. . .).
FT CARBOHYD 394 394 /FTid=CAR_000206.
FT FT 394 394 C-LINKED (MAN).
FT CARBOHYD 438 438 /FTid=CAR_000207.
FT FT 438 438 C-LINKED (MAN).
FT CARBOHYD 441 441 /FTid=CAR_000208.
FT FT 441 441 O-LINKED (FUC. . .).
FT CARBOHYD 450 450 /FTid=CAR_000209.
FT FT 450 450 C-LINKED (MAN).
FT CARBOHYD 498 498 /FTid=CAR_000210.
FT FT 498 498 O-LINKED (FUC. . .).
FT CARBOHYD 507 507 /FTid=CAR_000211.
FT FT 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1067 1067 T -> A (IN REF. 2, 3 AND 4).
FT FT 1067 1067 T -> A (IN REF. 2, 3 AND 4).
FT CONFLICT 84 84 T -> A (IN REF. 2).
FT FT 84 84 T -> A (IN REF. 2).
FT CONFLICT 523 523 1170 AA; 129412 MW; 69B3EDE5AE3A395E CRC64;
FT SEQUENCE 1170 AA; 129412 MW; 69B3EDE5AE3A395E CRC64;

Query Match 5.0%; Score 127; DB 1; Length 1170;
Best Local Similarity 44.4%; Pred. No. 0.035;
Matches 24; Conservative 7; Mismatches 13; Indels 10; Gaps 2;

QY 219 GEGDWSLWCVGCGNGKTRSC-----GYACT--ATESRTCDRNCP 262
Db 494 GWPSPMDICSVTCGGVQRRSLCNPPTQFGKDCVGDVTENQICRKQDCP 547

RESULT 11
AT52_BOVIN STANDARD; PRT; 1205 AA.
AC P79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI).
GN ADAMTS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=97225960; PubMed=9122202;
RA Colige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA Lapiere C.M.;
RA "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=95348096; PubMed=7622483;
RA Colige A., Beschin A., Sany B., Goebels Y., Van Beeumen J.,
RA Nusgens B.V., Lapiere C.M.;
RA "Characterization and partial amino acid sequencing of a 107-kDa
RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen.";

```

```

J. Biol. Chem. 270:16724-16730(1995).
-|- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -|- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC alpha-1(I) at Pro-I-Gln and of alpha-1(II) and alpha-2(I) chains
CC at Ala-I-Gln.
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -|- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVELS
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE N-TERMINUS IS BLOCKED.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -|- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -|- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X96389; CAA65253.1; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF01562; Rep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00090; tsp_1; 4.
DR SMART; SM00209; TSPI; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 253 BY SIMILARITY.
FT CHAIN 254 1205 ADAMTS-2.
FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 403 403 BY SIMILARITY.
FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 474 554 DISINTEGRIN-LIKE.
FT DOMAIN 555 611 TSP TYPE-1 1.
FT DOMAIN 612 716 CYS-RICH.
FT DOMAIN 717 845 SPACER.
FT DOMAIN 846 905 TSP TYPE-1 2.
FT DOMAIN 906 968 TSP TYPE-1 3.
FT DOMAIN 969 1024 TSP TYPE-1 4.
FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 31 35 POLY-ALA.
FT DOMAIN 177 180 POLY-GLU.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;

Query Match 4.9%; Score 124; DB 1; Length 1205;
Best Local Similarity 20.9%; Pred. No. 0.063;
Matches 64; Conservative 26; Mismatches 84; Indels 132; Gaps 15;

QY 192 PPRGWDHTAPGHRTE-----TKDPEYDSTDGEGDWSLWS---VCSVTCNGNGNKRTR 242
Db 526 PPLDGTWCAPGKHCFCGHCITWLPDILKRD-----GNWGAWSFGSCSTCGTGVKFRTR 580
QY 243 SC-----GYACT--ATESRTCDRNCPCGIEDTFTTAATEVSLLAGSEFNATKLFEV 292
Db 581 QCDNPFPANGRTCSGLAYDQLCNSQDCPDALDFR----- 617
QY 293 DTDSCERWMSCKSEFLKXKMKVNDLPSCSYPTVEVAYSTADIFDRIKRDPRW----- 348
Db 618 -EEOCRQW-----DLYPEHGDAQHHWLPHE 641
QY 349 -KDAAGPKKLEIY---KPTARYCIRSMLSLESTLLAAQHCYGDNNQLITRG--KGAGT 402
Db 642 HRDA---KERCHLYCESKETG-----EVSMMKRMVHDGTRCSYKDAFSLVCRGDKRVGC 693
QY 403 PNLISTEFSAE-----LHKYVDVLPWLICKGDMRSYNEARPPNNGOKCTESPSED 453
Db 694 DGVIGSSKQEDKCGVCGDGNHCKV-----VKGTSR-----SPKGLG 731
QY 454 YIKOFQ 459
Db 732 YIKMFE 737

RESULT 12
ATSA_HUMAN
AC ATSA4_HUMAN STANDARD; PRT; 837 AA.
ID 075173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
```

```
RA Trzaskos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RT family of proteins.";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
RL J. Biol. Chem. 275:25791-25797(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -!- INDUCTION: BY INTERLEUKIN-1.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: AB014588; BAA31663.1; -
EMBL: AF148213; AAD41494.1; -
EMBL: AY044847; AAL02262.1; -
HSSP: P34179; ILAG.
MEROPS: M12.221; -.
MIM: 603876; -.
InterPro: IPR001762; Disintegrin.
InterPro: IPR001590; Reprolysin.
InterPro: IPR000884; TSP1.
InterPro: IPR000130; Zn_MTpeptdse.
Pfam: PF01421; Reprolysin; 1.
Pfam: PF00090; tsp_1; 1.
SMART: SM00209; TSP1; 1.
PROSITE: PS00142; ZINC_PROTEASE; 1.
PROSITE: PS50092; TSP1; 1.
PROSITE: PS50092; ADAM_MEPRO; 1.
PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydroxylase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51 POTENTIAL.
FT PROPEP 52 212
FT CHAIN 213 837 ADAMTS-4.
FT SITE 194 194 CYSTEINE SWITCH (POTENTIAL).
FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 362 362 BY SIMILARITY.
FT METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
```



[illegible]



Db 385 DSEEGSPWSDWTKCSVTCSGTOMRSCDVTRSACTPHIQTRMCSFKKC 436

RESULT 15

AT54\_RAT

ID AT54\_RAT STANDARD; PRT; 630 AA.

AC Q9ESP7; Q9ESP8; Q9ESP6;

DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1) (Fragment).

DN ADAMTS4.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Brain;

RX MEDLINE=20415831; PubMed=10961658;

RA Satoh K.; Suzuki N.; Yokota H.;

RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin motifs) is transcriptionally induced in beta-amyloid treated rat astrocytes.";

RL Neurosci. Lett. 289:177-180(2000).

CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES.

CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393 site.

CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.

CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

CC -----

DR EMBL: AB042272; BAB16474.1; -

DR EMBL: AB042271; BAB16473.1; -

DR EMBL: AB042273; BAB16475.1; -

DR InterPro: IPR001590; Reprolysin.

DR InterPro: IPR000884; TSP1.

DR Pfam: PF00090; tsp.1; 1.

DR SMART: SM00209; TSP1.1.

DR PROSITE: PS50215; ADAM\_MEPRO; 1.

DR PROSITE: PS50092; TSP1.1.

DR PROSITE: PS00142; ZINC\_PROTEASE; 1.

DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE NEG.

KW Hydrolase; Metalloprotease; zinc; Glycoprotein; zymogen;

FT NON\_TER 1 1

FT PROPEP <1 5 BY SIMILARITY.

FT CHAIN 6 630 ADAMTS-4.

FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT\_SITE 155 155 BY SIMILARITY.

FT METAL 158 158 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 164 164 ZINC (CATALYTIC) (BY SIMILARITY).

FT DOMAIN 233 303 DISINTEGRIN-LIKE.

FT DOMAIN 316 367 TSP TYPE-1.

FT DOMAIN 368 478 CYS-RICH.

FT DOMAIN 479 630 SPACER.

FT DOMAIN 40 45 POLY-ALA.

FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 630 AA; 68384 MW; 63A428753167C7EF CRC64;

Query Match 4.7%; Score 118; DB 1; Length 630;

Best Local Similarity 21.9%; Pred. No. 0.082;

Matches 60; Conservative 19; Mismatches 91; Indels 104; Gaps 10;

QY 104 LLDLPNPDLSKADINGQXPNQIVTIEVVDGPDSEADKQDH-----PENKPSWSVPSDWR 159

Db 217 LLDKPEAP-----LHLPTVF---PKDYDADROCOLTEGPDSSHCPLPPCAA 262

QY 160 AWMORSLSARANGDQDYXYDSTSDSDSNFLNPPRGWDHTAPGHRHTETKDKQDEYDST-- 217

Db 263 LMCEGHLN-----GHAMCQTKHSPWADGTPC 288

QY 218 -----DGEGDWSLWSVCSVTGCGNGNOKRTRSC----- 244

Db 289 GPAQACMGGRCLHVDQLKDFNIPQAGGWPMPGDCSRCTCGGVQVSSRDCTKPVPRNG 348

QY 245 GYACTA--TESRTCDPRNCP-GIEDTFTATAATEVSLLAGSEEFNATKLFVEYDTSCEERW- 300

Db 349 GKYCEGRRTTPRSCTNKCPHGSALTER-----EEQCAAYNHRTDLFKSFPGCPMDWVPRYT 404

QY 301 -----MSCKSEFLKKYMHKVMNDLPSCPCS 325

Db 405 GVAPRDQCKLTQARALGYVYVLEPRVADGTPCS 438

Search completed: April 25, 2002, 17:27:30

Job time: 9602 sec

